Products

Encyclopedia of Forensic Sciences, Three-Volume Set, 1-3

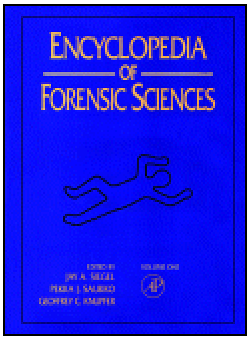
Book information
Product description
Audience
Author information and services

Ordering information
Bibliographic and ordering information
Conditions of sale

Book related information
Submit your book proposal
Product related news & events
Other books in same subject area

ENCYCLOPEDIA OF FORENSIC SCIENCES, THREE-VOLUME SET, 1-3

To order this title, and for more information, go to



http://books.elsevier.com/bookscat/links/details.asp?isbn=0122272153

Edited By
Jay Siegel , Michigan State University, East Lansing, U.S.A.
Geoffrey Knupfer , National Training Center for Scientific Support to Crime Investigation, Harperley Hall, Crook, UK
Pekka Saukko , University of Turku, Finland

Description
The **Encyclopedia of Forensic Sciences** is the first resource to provide comprehensive coverage of the core theories, methods, techniques, and applications employed by forensic scientists. One of the more pressing concerns in forensic science is the collection of evidence from the crime scene and its relevance to the forensic analysis carried out in the laboratory. The **Encyclopedia** will serve to inform both the crime scene worker and the laboratory worker of their protocols, procedures, and limitations. The more than 200 articles contained in the **Encyclopedia** form a repository of core information that will be of use to instructors, students, and professionals in the criminology, legal, and law enforcement communities.

Audience
Forensic science laboratories, police departments, academic libraries, law firms and law school libraries, academic departments teaching forensics, government agencies, and public libraries.

Contents
CONTENTS: Accident Investigation (a) Aircraft. Accident Investigation (b) Motor vehicle (including biomechanics of injuries). Accident Investigation (c) Rail. Accident Investigation (d) Reconstruction. Accident Investigation (e) Airbag related injuries and deaths. Accident Investigation (f) Determination of cause. Accident Investigation (g) Driver versus passenger in motor vehicle collisions. Accident Investigation (h) Tachographs. Accreditation of Forensic Science Laboratories. Administration of Forensic Science (a) An international perspective. Administration of Forensic Science (b) Organisation of laboratories. Alcohol (a) Blood. Alcohol (b) Body fluids. Alcohol (c)

Breath. Alcohol (d) Post-mortem. Alcohol (e) Interpretation. @con:CONTENTS: Alcohol (f) Congener analysis. Analytical Techniques (a) Separation techniques. Analytical Techniques (b) Microscopy. Analytical Techniques (c) Spectroscopy. Analytical Techniques (d) Mass spectrometry. Anthropology: Archaeology. Anthropology: Skeletal Analysis (a) Overview. Anthropology: Skeletal Analysis (b) Morphological age estimation. Anthropology: Skeletal Analysis (c) Sex determination. Anthropology: Skeletal Analysis (d) Determination of racial affinity. Anthropology: Skeletal Analysis (e) Excavation/retrieval of forensic remains. Anthropology: Skeletal Analysis (f) Bone pathology and ante-mortem trauma in forensic cases. Anthropology: Skeletal Analysis (g) Skeletal trauma. Anthropology: Skeletal Analysis (h) Animal effects on human remains. Anthropology: Skeletal Analysis (i) Assessment of occupational stress. Anthropology: Skeletal Analysis (j) Stature estimation from the skeleton. Art and Antique Forgery and Fraud. Autoerotic Death. Basic Principles of Forensic Science. Biochemical Analysis (a) Capillary electrophoresis in forensic science. Biochemical Analysis (b) Capillary electrophoresis in forensic biology. Blood Identification. Blood Stain Pattern Analysis and Interpretation. Causes of Death (a) Post-mortem changes. Causes of Death (b) Sudden natural death. Causes of Death (c) Blunt injury. Causes of Death (d) Sharp injury. Causes of Death (e) Gunshot wounds. Causes of Death (f) Asphyctic deaths. Causes of Death (g) Burns and scalds. Causes of Death (h) Traffic deaths. Causes of Death (i) Systemic response to trauma. Causes of Death (j) Poisonings. Cheiloscopy. Clinical Forensic Medicine (a) Overview. Clinical Forensic Medicine (b) Defence wounds. Clinical Forensic Medicine (c) Self-inflicted injury. Clinical Forensic Medicine (d) Child abuse. Clinical Forensic Medicine (e) Sexual assault and semen persistence. Clinical Forensic Medicine (f) Evaluation of gunshot wounds. Clinical Forensic Medicine (g) Recognition of pattern injuries in domestic violence victims. Computer Crime. Credit Cards: Forgery and Fraud. Crime-Scene Investigation and Examination (a) Recording. Crime-Scene Investigation and Examination (b) Collection and chain of evidence. Crime-Scene Investigation and Examination (c) Recovery. Crime-Scene Investigation and Examination (d) Packaging. Crime-Scene Investigation and Examination (e) Preservation. Crime-Scene Investigation and Examination (f) Contamination. Crime-Scene Investigation and Examination (g) Fingerprints. Crime-Scene Investigation and Examination (h) Suspicious deaths. Crime-Scene Investigation and Examination (i) Major incident scene management. Crime-Scene Investigation and Examination (j) Serial and series crimes. Crime-Scene Investigation and Examination (k) Scene analysis/reconstruction. Crime-Scene Investigation and Examination (l) Criminal analysis. Crime-Scene Investigation and Examination (m) Decomposing and skeletonized cases. Criminal Profiling. Criminalistics. Detection of Deception. Disaster Victim Identification. DNA (a) Basic principles. DNA (b) RFLP. DNA (c) PCR. DNA (d) PCR-STR. DNA (e) Future analytical techniques. DNA (f) Paternity testing. DNA (g) Significance. DNA (h) Mitochondrial. Document Analysis (a) Handwriting. Document Analysis (b) Analytical methods. Document Analysis (c) Forgery and counterfeits. Document Analysis (d) Ink analysis. Document Analysis (e) Printer types. Document Analysis (f) Document dating. Drugs of Abuse (a) Blood. Drugs of Abuse (b) Body fluids. Drugs of Abuse (c) Ante-mortem. Drugs of Abuse (d) Post-mortem. Drugs of Abuse (e) Drugs and driving. Drugs of Abuse (f) Urine. Drugs of Abuse (g) Hair. Drugs of Abuse (h) Methods of analysis. Drugs of Abuse (i) Designer drugs. Dust. Ear Prints. Education, An International Perspective. Electronic Communication and Information. Entomology. Ethics. Evidence (a) Classification of evidence. Evidence (b)The philosophy of sequential analysis. Evidence (c) Statistical interpretation of evidence/Bayesian analysis. Expert Witnesses, Qualifications and Testimony. Explosives, Methods of Analysis. Facial Identification (a) The lineup, mugshot search and composite. Facial Identification (b) Photo image identification. Facial Identification (c) Computerized facial reconstruction. Facial Identification (d) Skull-photo superimposition. Facial Identification (e)

Facial tissue thickness in facial reconstruction. Fibres (a) Types. Fibres (b) Transfer and persistence. Fibres (c) Recovery. Fibres (d) Identification and comparison. Fibres (e) Significance. Fingerprints (Dactyloscopy) (a) Visualisation. Fingerprints (Dactyloscopy) (b) Sequential treatment and enhancement. Fingerprints (Dactyloscopy) (c) Identification and classification. Fingerprints (Dactyloscopy) (d) Standards of proof. Fingerprints (Dactyloscopy) (e) Chemistry of print residue. Fire Investigation (a) Types of fire. Fire Investigation (b) Physics/Thermodynamics. Fire Investigation (c) Chemistry of fire. Fire Investigation (d) The fire scene. Fire Investigation (e) Evidence recovery. Fire Investigation (f) Fire scene patterns. Fire Investigation (g) The laboratory. Firearms (a) Types of weapons and ammunitions. Firearms (b) Range and penetration. Firearms (c) CS Gas. Firearms (df) Humane killing tools. Firearms (e) Laboratory analysis. Forensic Anthropology. Forensic Engineering. Forensic Nursing. Forensic Psycholinguistics. Forensic Toxicology (a) Overview. Forensic Toxicology (b) Methods of analysis - ante-mortem. Forensic Toxicology (c) Methods of analysis - post-mortem. Forensic Toxicology (d) Interpretation of results. Forensic Toxicology (e) Inhalants. Forensic Toxicology (f) Equine drug testing. Forgery and Fraud (a) Overview (including counterfeit currency). Forgery and Fraud (b) Auditing and accountancy. Gas Chromatography, Methodology in Forensic Sciences. Genetics (a) Serology. Genetics (b) DNA - statistical probability. Glass. Hair (a) Background. Hair (b) Hair transfer, persistence and recovery. Hair (c) Identification of human and animal hair. Hair (d) Microscopic comparison. Hair (e) Other comparison methods. Hair (f) Significance of hair evidence. Hair (g) DNA typing. Health and Safety (including Risk Assessment). History (a) Crime scene sciences. History (b) Fingerprint sciences. Identification/Individualization, Overview and Meaning. Investigative Psychology. Legal Aspects of Forensic Science. Lie Detection (Polygraph). Literature and the Forensic Sciences (a) Resources. Literature and the Forensic Sciences (b) Fiction. Microchemistry. Modus Operandi. Odontology. Offender Signature. Paints and Coatings: Commercial, Domestic and Automotive. Pathology (a) Overview. Pathology (b) Victim recovery. Pathology (c) Autopsy. Pathology (d) Preservation of evidence. Pathology (e) Post-mortem changes. Pathology (f) Post-mortem interval. Pattern Evidence (a) Footmarks (footwear). Pattern Evidence (b) Footmarks (bare footprints). Pattern Evidence (c) Shotgun ammunition on a target. Pattern Evidence (d) Tools. Pattern Evidence (e) Plastic bag striations. Pattern Evidence (f) Serial number. Pharmacology. Post-Mortem Examination, Procedures and Standards. Psychological Autopsies. Psychology and Psychiatry (a) Overview. Psychology and Psychiatry (b) Psychiatry. Psychology and Psychiatry (c) Psychology. Quality Assurance/Control. Serial Killing. Soil and Geology. Stalking. Statistical Interpretation of Evidence. Time Factor Analysis. Voice Analysis. Wildlife. Wood Analysis.

INTRODUCTION

THERE ARE those who say that there really is no such thing as 'forensic science'; that instead, it is a collection of scientific techniques and principles that are begged and borrowed from 'real' sciences such as chemistry, biology, physics, medicine and mathematics. Others have suggested that the 'sciences' of fingerprints, firearms and toolmarks and questioned documents are the only real forensic sciences, and all the rest of it is on loan from the classical hard sciences. There is some truth to both of these definitions. Few would doubt that much of the process of forensic DNA typing consists of classical laboratory methods of DNA characterization, applied to real world evidentiary material, much of which has been degraded or contaminated. Or that in the case of controlled substance analysis, the methods of analytical chemistry are applied to the characterization and comparison of so-called 'street' drugs.

However, although many of the techniques used in modern forensic science have been borrowed from other sciences, it is also true that in recent years, it has matured into a scientific discipline in its own right. Many scientific techniques used in the analysis of physical evidence have been developed and perfected principally for forensic purposes. Also, forensic science does not just involve analysis of chemical, physical and biological materials. There are important considerations of collection and preservation of evidence, interpretation of findings from analysis, and presentation of expert, scientific testimony in criminal and civil courts. These processes are not isolated, but in fact must be carefully integrated, if forensic science is to have any meaningful effect on the criminal and civil justice systems in the world. This is what makes forensic science a unique and fascinating field of study and work.

The *Encyclopedia of Forensic Sciences* was conceived and developed with these principles in mind. It brings together, for the first time in such breadth and depth, a collection of articles that cover the whole panoply of activities which make up forensic science. Although the main focus of the Encyclopedia is the analysis of evidence, considerable space is devoted to investigative aspects including collection and preservation of evidence, and the interpretation of analytical findings and their presentation in court. The editors want the reader to be mindful of the interconnections between the various disciplines and factors that comprise the forensic science enterprise, all of which are necessary to make it work. We also want the reader to appreciate that, although there are differences in how various agencies worldwide handle and analyze evidence – and how expert testimony is offered in court, there is the same basic framework and interdependence at work everywhere.

We hope and trust that the reader will see the Encyclopedia as a valuable resource in navigating the interesting and ever-evolving world of forensic science.

A

ACCIDENT INVESTIGATION

Contents

Airbag Related Injuries and Deaths
Determination of Cause: Overview
Determination of Cause: Reconstruction
Driver Versus Passenger in Motor Vehicle Collisions
Motor Vehicle
Rail
Tachographs

Airbag Related Injuries and Deaths

W S Smock, University of Louisville, Louisville, KY, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0748

Introduction

During the past decade the motoring public has been shocked to learn that air bags, a life-saving device promoted by the automotive industry, can also induce severe and fatal injuries. Over the last 10 years in the United States, nearly 200 men, women and children have been fatally injured by deploying air bags. Thousands more have sustained serious nonfatal injuries, including cervical spine fractures, closed head injuries, and multiple fractures and amputations of digits and hands. Ironically, the vast majority of these serious and fatal injuries were incurred in low and moderate speed collisions in which little or no injury would have been otherwise expected.

Historical Context

The first air bag patents were filed in 1952. Ford and General Motors began experimenting with these early prototypes in the late 1950s. Based on documents from the automotive industry, it was apparent, even as early as 1962, that deploying air bags had the potential to induce serious and fatal injuries, particularly to children. These early documents include analyses of tests conducted by Ford at their automotive safety research office in the late 1960s. The tests demonstrated that there was sufficient force associated with air bag deployment to traumatically eject a child from a vehicle. Ford also noted the amputation of a steel-hinged arm from a dummy secondary to the explosive force of deployment. When testing involved the use of animal models, the list of severe and fatal injuries grew. Cardiac rupture, hepatic rupture, splenic rupture, aortic and vena cava transection, atlanto-occipital dislocation, cervical spine fractures, severe closed head injury and decapitation were observed.

Testing in the 1980s by the automotive manufacturers continued to demonstrate the risk of injury induction. One study conducted and reported by General Motors indicated 'many of the exposures were at loading severities beyond the level representing an estimate of nearly 100% risk of severe injury'. These laboratory tests were completed and the knowledge available to the entire automotive industry well *before* air bags were placed in American-made vehicles.

With 40 years of research behind us and all of the resultant data before us, it is apparent that steering wheel and dash-mounted air bags, devices designed to protect occupants in high-speed frontal collisions, can also maim and kill.

Automotive Industry

During the period of 1960 to 1990, the automobile industry embraced an extensive research and development program regarding air bags. This testing involved anthropomorphic dummies, anesthetized swine and baboons, and even human volunteers. It was recognized early on that occupants in the path of the deploying air bag, or who were positioned close to the air bag at the moment of deployment, were at a very significant risk of receiving a severe or fatal injury. In January 1970, at a meeting of engineers from General Motors, Ford and Chrysler, this information was discussed. As a result, a Chrysler engineer wrote: 'This is a very serious problem that must be resolved for the inflatable restraint. Having a child directly in front of the bag when it inflates could prove fatal.'

General Motors tested air bags on baboons at Wayne State University during the early 1970s. The research indicated that 'if the head is in the path of the deploying air bag, it is concluded that injury is likely to occur in the form of brain or neck injury to a child'.

Testing by Ford in the 1970s revealed that individuals involved in collisions at less than 32 k.p.h. (20 m.p.h.) experienced more severe injuries and loading forces in vehicles equipped *with* an air bag than those without one. Ford also noted that there was 'overwhelming evidence that air bags may be dangerous for small children'. In 1972, a Ford engineer wrote the following warning, which was never placed in a vehicle: 'The right front seat should be used only by persons who are more than five feet [1.52 m] tall and are in sound health. Smaller persons and those who are aged or infirm, should be seated and belted in the rear seat.'

In a series of tests conducted by General Motors in the 1980s, anesthetized swine were placed with their chests in close proximity to the air bag module. The tests revealed that when the swine's thorax was impacted by the force of the deploying air bag and the air bag module, significant thoracic and abdominal injuries were sustained. These injuries in one case included: 17 rib fractures; two cardiac perforations; a splenic laceration; a liver hematoma and death within 30 min. The proximity of the occupant to the deploying bag and module cover were the pivotal contributory factors.

Human Injuries

The serious and life-threatening injuries that were originally observed in the industry's laboratories using animal models began to be observed in humans on US roadways in the 1990s. The first six driver air bag-related deaths which were investigated by the government and the automotive industry revealed that the majority of these victims were women of short stature. It was also noted that the fatal injuries could be incurred even if the occupant was restrained by a lap and chest belt. The injuries initially seen included massive head injuries with diffuse axonal injury, subdural hematomas and skull fractures. Additional injuries evaluated included cervical spine fracture, cardiac perforation, pulmonary contusions and multiple rib fractures.

Sodium azide is the explosive propellant used to initiate the deployment cycle in most air bag designs in use today (**Fig. 1**). When sodium azide is ignited, the deploying air bag explodes toward the occupant at speeds of up to 336 k.p.h. (210 m.p.h.). An air bag system has two components, either one of which may induce injuries: the canvas-covered air bag itself and the air bag module cover (**Fig. 2**). Injuries incurred during deployment are relevant to the component inflicting them.

Obviously, the types of injuries which result from impact with the canvas air bag are different from

Figure 1 The air bag is transported as an explosive material. Sodium azide is the explosive propellant used in the majority of air bag modules.

Figure 2 The injury-producing components of the air bag system are the air bag and the module cover which overlies it.

those which result from impact with its module cover. There are three phases to air bag deployment: 'punch out', 'catapult' and 'bag slap'. Injuries can be inflicted at any point during the deployment process:

- *Punch out* This is the initial stage of deployment. If the bag makes contact at this stage, the following injuries can result: atlanto-occipital dislocation, cervical spine fracture with brainstem transection, cardiac, liver and splenic lacerations, diffuse axonal injuries, subdural and epidural hematomas, and decapitation.
- *Catapult* This is the midstage of deployment when rapidly inflating bag 'catapults' or drives the head and neck rearward. This occurs with sufficient energy to rupture blood vessels, ligaments and fracture cervical vertebrae. The neck injuries occur as the result of cervical spine hyperextension.
- *Bag slap* This is the final stage of deployment which occurs at the bag's peak excursion. Appropriately named, this happens when the canvas bag's fabric may 'slap' the occupant's face, resulting in injuries to the eye and epithelium.

The air bag module covers are located in the steering wheel on the driver's side and in the dashboard panel on the passenger side. As the bag deploys, the module cover is also propelled outward at speeds of up to 336 k.p.h. (210 m.p.h.). Most steering wheel designs house the horn within the air bag module compartment. Hand and arm injuries observed in individuals whose extremities were in contact with the module at the moment of its rupture include: degloving, fracture dislocation, fracture dislocation and amputations (partial and complete of digits and forearms). If the module cover makes contact with an occupant's face, head or neck, skull fractures and severe or fatal head injuries, and decapitations have also been observed. The driver's side cover is generally made with a rubberized plastic type of material, while the passenger side may have a metal housing. Contact with either type can prove fatal.

Specific Injury Patterns

Ocular

The eye is extremely vulnerable to air bag-induced injury. These injuries range from corneal abrasions from contact with the air bag, chemical burns from contact with unburned sodium azide, to retinal detachment and globe rupture from the blunt force trauma of the expanding bag (**Fig. 3**). The wearing of eyeglasses in some cases has proven to be of benefit, as it offers a degree of barrier protection between the eye and the deploying bag.

Figure 3 (see color plate 1) This patient sustained a severe corneal abrasion secondary to the membrane forces associated with air bag deployment.

Face and head

The most common injury associated with air bag deployment is that of facial abrasion. The abrasions result from a sliding contact between the bag and the face (**Fig. 4**). The injuries are not chemical 'burns' but deep abrasions.

Figure 4 Abrasions to the cheeks, forehead and nose are the most common injury associated with air bag deployment.

Cranial and intracranial

When acceleration forces are applied to the cranial vault, a variety of traumatic injuries to the brain and surrounding structures will result. These include subdural hematomas, cortical contusions, atlanto-occipital dislocations, skull fractures and brainstem transections. Cranial injuries may result from contact with either the deploying of bag or module cover (**Fig. 5**).

Cervical spine

The blow from an air bag or module cover which produces a rapid and violent hyperextension of the cervical spine of the driver or passenger will have significant consequences for the cervical vertebrae. Injuries commonly seen as a result of hyperextension include atlanto-occipital dislocation, comminuted fractures of one or more upper cervical vertebrae, rupture of the anterior and posterior longitudinal spinal ligaments, and cervical spine disarticulation with transection of the cervical cord. The majority of these injuries are associated with the upper cervical vertebrae, although lower cervical vertebrae injuries have been observed.

Extremities

The upper extremities are very vulnerable to traumatic injury from the deploying bag and its module cover. When an individual's hand or forearm is on or near the module cover at the moment of deployment, the occupant can expect to sustain multiple fractures, and/or tissue degloving or amputation of fingers, hand or forearm (**Fig. 6**). The horn-button-within-the-module cover design significantly increases the risk of injury to the occupant's upper extremities at the moment of deployment. Many of these upper extremity injuries are associated with an occupant's attempt to blow the horn, the button of which is located within the module cover. Forces from air bag deployment may be transmitted to the hand, wrist or forearm and may even involve the humerus. It is not unusual to see significantly comminuted fractures involving the wrist, forearm, elbow and distal humerus (**Figs 7** and **8**). The vehicles whose module

Figure 5 Cranial or facial contact with the deploying bag or

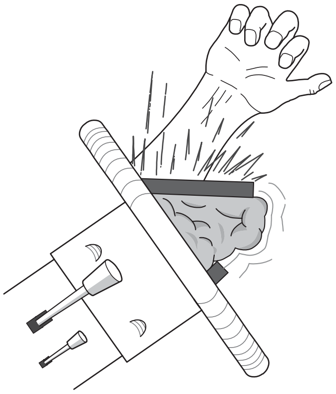
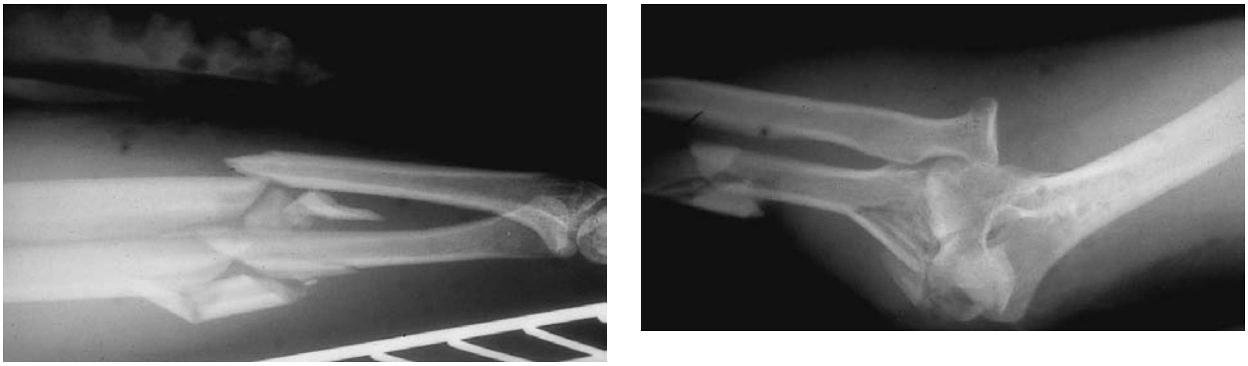


Figure 6 When the forearm is located in a horizontal fashion



A **B**

Figure 7 When the hand, wrist or forearm is on or near the module cover at the moment of deployment, for example when blowing the horn, significant fractures, degloving injuries and amputations will result. **(A)** This open-comminuted bending fracture of the radius and ulna was the result of contact with the module cover. **(B)** This patient sustained a comminuted fracture of the proximal and midshaft ulna as well as a radial head dislocation from impact with the module cover.

A

B

Figure 8(A) and (B) These severely comminuted fractures resulted when this individual was blowing her horn. She sustained a frontal impact of approximately 15 k.p.h., which resulted in air bag activation. The horn activation button was located within the module cover, which explains why the patient had her forearms over the module cover at the moment of deployment.

covers are of a higher mass have the propensity to inflict more severe injuries. Some of the worst offenders are module covers located on the passenger side, which may have a soft coating of plastic on the exterior but have an underlying piece of rigid metal (**Fig. 9B**). The placement of hands on the passenger-side dashboard, in a bracing maneuver, has resulted in the traumatic amputation of hands and forearms (Fig. 9A).

Respiratory

The byproducts of combustion as well as other inert materials within the air bag may produce a white cloud within the vehicle. Many occupants have thought that this indicated a vehicle fire. This whitish material is principally cornstarch, talc and the byproducts of sodium azide combustion. There may be a small percentage of unburned sodium azide present within this powder as well. Inhalation of these materials can result in a chemical pneumonitis and the induction of asthma-type symptoms. These byproducts may also cause a chemical irritation of open wounds and a basic (high pH) burn to the eyes.

Sample Cases

Case 1

A 35-year-old female, 1.57 m (5′ 2″) and 50 kg (110 lb.), was the restrained driver in a 1991 Ford Taurus (**Fig. 10**). The vehicle's front bumper grazed a guard rail, which resulted in deployment of the driver's air bag. The patient sustained an immediate respiratory arrest, with subsequent declaration of brain death 12 h later.

A postmortem examination was conducted and revealed the following injuries: significant midface trauma with bilateral epistaxis; corneal abrasions; contusions of the chest; left subdural hematoma (overlying the frontal and parietal regions); subarachnoid hemorrhage and severe cerebral edema.

Examination of the exterior of the vehicle revealed very minor damage which was limited to the front

A

B

Figure 9 (A) This partially-amputated wrist was the result of placement of the front seat passenger's hand on the dashboard at the moment of deployment. **(B)** The underlying structure of the module cover was metal.

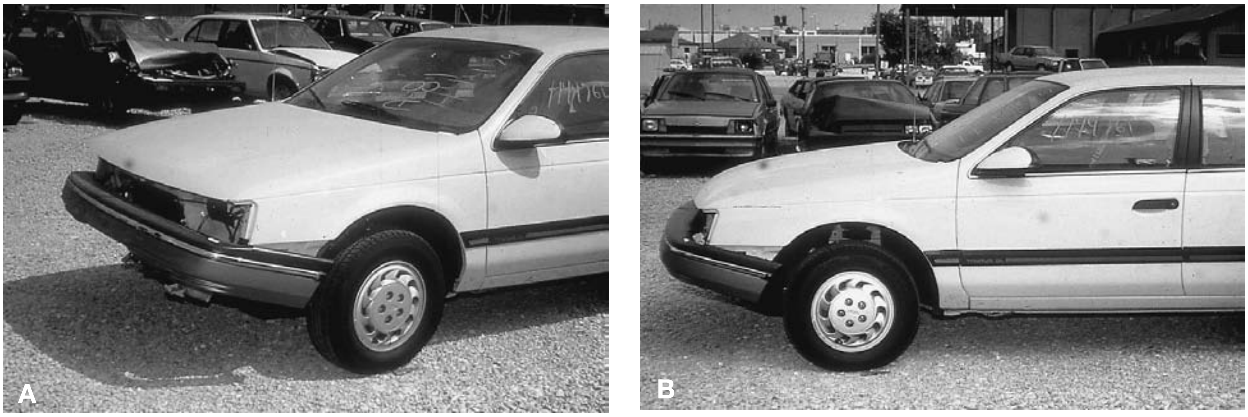


Figure 10(A) and (B) A 1991 Ford Taurus was involved in a very minor glancing collision between the front bumper and a guardrail. There was no underlying frame, fender or structural damage.

bumper. The examination of the interior revealed a tear on the left lower portion of the module cover (**Fig. 11**). This tear was the result of contact between the left side of the victim's face and the deploying module (Figs 5 and 11).

Case 2

A 4-year-old male was the lap belt-restrained occupant of the front right seat of a 1995 GEO Metro (**Fig. 12**). The shoulder portion of the harness was placed underneath the right arm. The patient was catapulted from the front passenger seat to the rear of the vehicle. The patient was found pulseless and apneic on arrival of emergency medical services.

A postmortem examination was conducted and revealed the following injuries: an atlanto-occipital dislocation with brainstem transection; large subdural hemorrhages; multiple rib fractures with underlying pulmonary contusions; liver and splenic lacerations; clavicular fracture and significant facial abrasions underlying the mandible bilaterally and on the right cheek (**Fig. 13A**). Examination of the abdomen also revealed a lap belt contusion below the umbilicus (Fig. 13B).

Examination of the vehicle revealed front end damage consistent with a change of velocity of less than 15 k.p.h. (9 m.p.h.). There was no damage to the driver or passenger compartments. Examination of the passenger bag revealed the presence of blood and tissue transfer. The patient's injuries resulted from blunt force trauma to the chest and abdomen as well as a hyperextension injury of the neck with a rapid rearward acceleration.

Case 3

A 35-year-old female was the front seat passenger in a 1995 Nissan Altima. The patient was in a lap–shoulder belt, with the passenger seat in the most rearward position. The vehicle was stopped in a line of traffic, waiting for a traffic signal to turn, when it was hit from behind by a vehicle travelling at the speed of approximately 5 k.p.h. (3 m.p.h.). The rear impact pushed the Nissan forward into the trailer hitch of the truck in front. This resulted in air bag

A B

Figure 11 (A) The left lower corner of the module cover exhibited tearing from contact with the left side of the driver's face. **(B)** Close-up of the module cover reveals the presence of a torn area. There is also an area which indicates this piece was removed by a sharp implement.

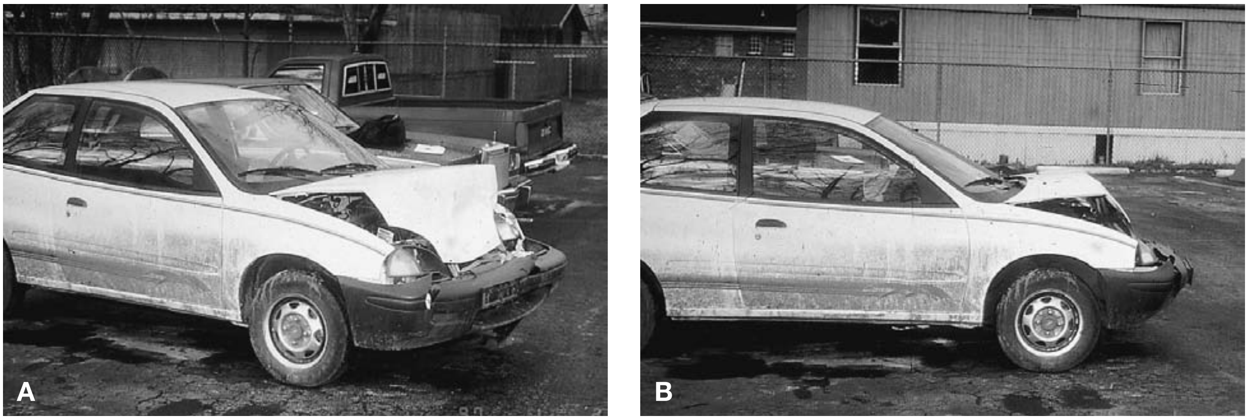


Figure 12 (A) and (B) A 1995 GEO Metro with minor front-end damage consistent with an impact speed of less than 15 k.p.h. (9 m.p.h.).

deployment. The air bag impacted the patient's left eye. Examination of the face revealed significant periorbital trauma. There were abrasions on the forehead as well as on the cheek and chin. Examination of the eye revealed the presence of chemosis, a hyphema and a retinal detachment (**Fig. 14**).

Examination of the vehicle revealed a 5×5 cm dent in the right front bumper (**Fig. 15**). There was no

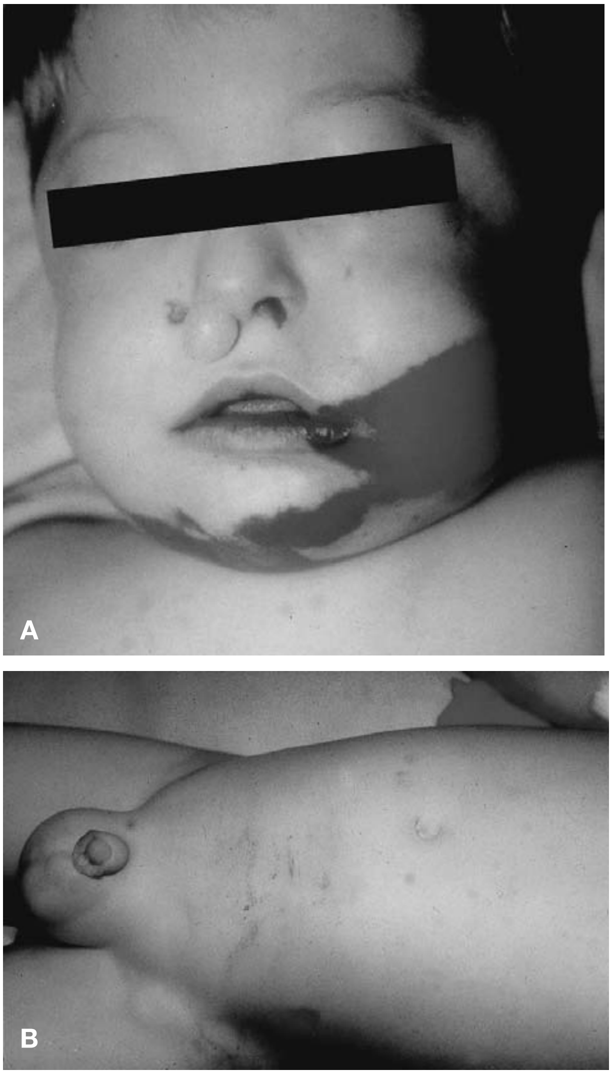


Figure 13 (A) This fatally injured 4-year-old male exhibits significant facial abrasion, overlying the mandible as well as abrasion on the left cheek. This was the result from contacts with the upwardly-deploying air bag. **(B)** Examination of the patient's abdomen reveals a horizontally-oriented contusion consistent with a lap belt.

Figure 14 This 1.72 m (5′ 8″) restrained patient suffered permanent retinal detachment secondary to air bag contact.

Figure 15(A) and (B) A 1995 Nissan Altima. The only damage to the vehicle is a 5 × 5 cm indentation in the front bumper from impact with the ball of a trailer hitch.

significant rear end damage. Examination of the bag revealed transfer of make-up and blood to the bag.

Forensics of Air Bag Injuries

Locard's principle regarding the transfer of physical evidence between two impacting objects is dramatically evidenced in the case of air bags and air bag-induced injuries. The transfer of evidence to the air bag itself may take various forms. Blood and epithelial tissue transfer is, of course, common but transfer of make-up, including lipstick, rouge and mascara, to the air bag is also seen (**Fig. 16**). Analysis of the blood spatter pattern on the bag may assist the investigator in determining the position of the occupant and the configuration of the steering wheel at the moment of air bag deployment.

Examination of the air bag module cover may reveal the presence of trace evidence. Depending on the design of the module cover, there may actually be tearing or bending of the module cover, indicative of contact with an occupant's more rigid (bony) surface: face or forearm. Scuff-type marks on the module cover indicate contact with an object, frequently the forearm (**Fig. 17**). Fabric imprints may also be seen on close inspection (**Table 1**).

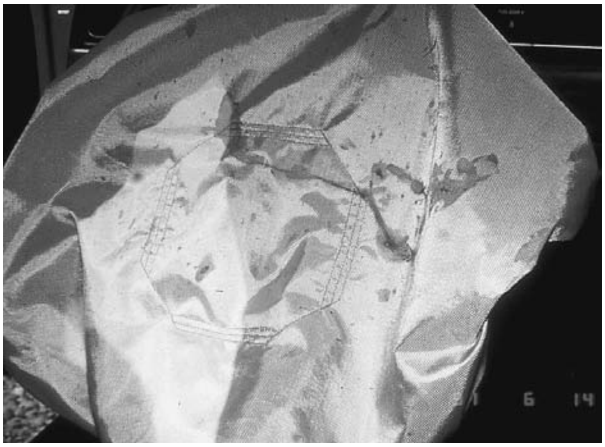


Figure 16 Close examination of an air bag may reveal a multitude of transferred evidence. This evidence will include: hair, blood, epithelial tissue and facial makeup.

Summary

In the United States the deploying air bag has been responsible for nearly 200 deaths and thousands of

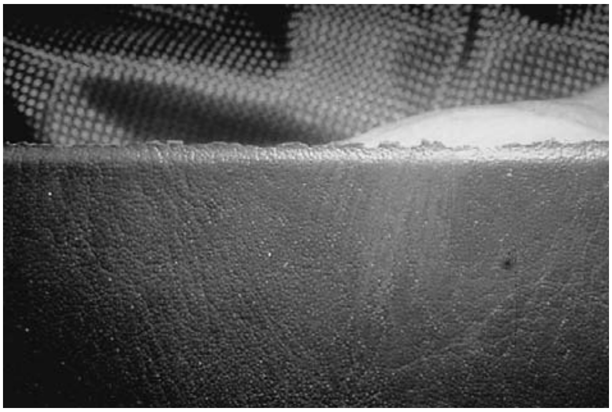


Figure 17 Close inspection of a module cover may reveal the presence of scuffing, and fabric imprints.

Table 1 Trace evidence.

Air bag
Blood
Make-up
Hair
Tissue
Module cover
Tearing of cover material
Scuff marking
Fabric imprint
Tissue
Hair

severe injuries. It is clear that the forces of deployment are not insignificant and must be respected by the vehicle occupant. A review of the literature indicates that the serious and fatal injuries, which were once produced in the laboratory setting, are now being observed in 'real-world' collisions. Many clinicians may not be aware of the injury-producing power of the deploying air bag and must be informed of the patterns associated with air bag-induced injuries. The motoring public must also be informed and warned of the dangers of air bag deployment, just as the automotive industry was over 30 years ago.

See also: **Accident Investigation**: Motor Vehicle.

Further Reading

Horsch J, Lau I and Andrzegkad *et al.* (1990) *Assessment of Air Bag Deployment Loads*, SAE 902324. Detroit: Society of Automotive Engineers.

Huelke DF, Moore JL, Compton TW *et al.* (1994) *Upper Extremity Injuries Related to Air Bag Deployments*, SAE 940716. Detroit: Society of Automotive Engineers.

Huelke DF, Moore JL and Ostrom M (1992) Air bag injuries and accident protection. *Journal of Trauma* 33:894–897.

Landcaster GI, Defrance JH and Borrusso J (1993) Air bag associated rupture of the right atrium. *New England Journal of Medicine* **328**:358.

Lau I, Horsch JD, Viano DC *et al.* (1993) Mechanism of injury from air bag deployment loads. *Accident Analysis and Prevention* **25**:29.

Mertz HJ and Weber DA (1982) *Interpretations of Impact Response to a Three Year Old Child Dummy Relative to Child Injury Potential*, SAE 826048. Detroit: Society of Automotive Engineers.

Mertz HJ, Driscoll GP, Lennox JB *et al.* (1982) Response of animals exposed to deployment of various passenger and portable restraint system concepts for a variety of collision severities and animal positions. *National Highway and Transportation Safety Administration. 9th International Technical Conference on Experimental Safety Vehicles*, pp. 352–368. Kyoto, Japan, November 1982.

Parents for Safer Air Bags (1997) *The Air Bag Crisis: Causes and Solutions*.

Prasad P and Daniel RP (1984) *A Biomechanical Analysis of Head, Neck and Torso Injuries to Child Surrogates Due to Sudden Torso Acceleration*. SAE 841656. Detroit: Society of Automotive Engineers.

Smock WS (1992) Traumatic avulsion of the first digit, secondary to air bag deployment. *Proceedings 36. Association for the Advancement of Automotive Medicine*, p. 444. Des Plains, IL.

Smock WS and Nichols GN (1995) Air bag module cover injures. *Journal of Trauma, Injury, Infection and Critical Care* 38:489–493.

Viano DC and Warner CV (1976) *Thoracic Impact Response of Live Porcine Subjects*, CvSAE 760823, 733–765. Detroit: Society for Automotive Engineers.

Determination of Cause: Overview

D Rudram, The Forensic Science Service, London, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0398

Introduction

The purpose of all accident investigation is to establish the 'cause' of the incident. This information will be sought for a number of reasons which may overlap. Thus the victims or their family will want to know why they were injured or killed, regulatory authorities may wish to establish responsibility and take appropriate action as a result and safety authorities will want to see what can be done to prevent a re-occurrence.

What do we mean by the cause of an accident and if it was caused by the actions or omissions of someone is it really an accident at all? The term 'dangerous occurrence' used in Health and Safety Regulations is in many ways more appropriate but the neutral term 'incident' will be used here. There can be several causes for one of these dangerous occurrences and the causes are often multilayered; the investigator needs to keep this firmly in mind when analyzing the data and reaching a conclusion. As always the investigator needs to have a clear understanding of the available evidence and what it means. An example from some years ago illustrates both of these points.

In 1975 a train on the London Underground ran into the blank wall at the end of the line at Moorgate Station. The immediate cause of the incident was the failure of the train to stop at the final station which was brought about because the driver did not obey the signals which controlled the speed of approach of the train. But why did the driver ignore these signals? There was alcohol in both the driver's blood and the drink that he had with him when they were analyzed post mortem. An explanation for his actions perhaps? The driver, however, was an abstemious man and the incident happened early in the morning rush hour so, although not impossible, alcohol seemed an implausible explanation. Also as the incident occurred underground, the driver's body was not recovered for several days and the temperature in the tunnel was elevated, the conclusion that postmortem putrefaction rather than antemortem drinking was the explanation for the presence of alcohol becomes more likely. This incident illustrates that sometimes the ultimate cause of a dangerous occurrence is never known.

Sources of information

Establishing the sequence of events which led up to an incident is usually an essential prerequisite for establishing why the incident happened. To do so successfully the investigator needs information and usually the more information the better. Although this can lead to problems of assimilation, arbitrarily excluding some sources or types of information runs the risk of excluding the vital fact which enables an investigator to unravel the puzzle. Information can be obtained from many sources each with its own advantages and disadvantages all of which need to play a part in the interpretation of events.

Eye witnesses

The evidence of witnesses needs to be treated with great care. People can genuinely believe they saw something which was impossible but this should not cause the investigator to dismiss them out of hand. Some of their observations may be crucial. An eye witness can provide information that is available from no other source. For example:

- Did the pedestrian walk quickly or slowly?
- Did they cross the road in a straight line?
- What color were the traffic signals?

Questions that the investigator should be asking about this evidence should include:

- Where was the witness and what could he or she see from there?
- Is their evidence supported by another witness or other information?
- Are the stories of two witnesses too similar, suggesting that they have colluded, albeit innocently?

Finally it should be remembered that some research suggests that certainty is no guarantee of accuracy of recall.

Tire and other marks in the road

One of the most common sources of information in the investigation of road traffic incidents are tire marks, scrapes and gouges left by the vehicles in the road before, during and after the impact. These marks can provide both qualitative and quantitative information about the incident. The position and orientation of tire marks indicate the path of the vehicles before and after the incident and changes in the direction of tire marks, gouges and scrapes in the road surface can indicate the position of the vehicle at impact (**Fig. 1**). The nature and dimensions of the tire marks may enable the investigator to calculate the speed of one or more of the vehicles which is usually an essential component of any analysis of the cause of the incident.

Figure 1 Tire marks scrapes and gouges can be seen in the roadway following this incident involving two goods vehicles. © Metropolitan Police Service.

The investigator needs to understand how the marks were made and as a consequence what that tells them about the incident. Speed estimates are an important component in any analysis of road traffic incidents but care needs to be used in interpreting them. The accident investigator needs to consider how precise is the estimate and at what point in the developing situation was the vehicle traveling at that speed.

Vehicle damage

The damage sustained by the vehicles also provides both qualitative and quantitative information. The nature and position of the damage will indicate the direction of the applied force from which the relative movements of the vehicle or vehicles can be deduced (**Fig. 2**). If one of the vehicles is stationary the problem is simplified considerably. The extent of the damage can be used to calculate the velocity change caused by the impact. Damage alone does not enable the impact speed to be calculated without some additional information. Again impact with a stationary object is the simplest case to deal with but in other circumstances the post impact movements of the vehicles need to be considered.

It is very easy to misinterpret vehicle damage and the investigator needs to understand the limitations of the model that is being used. The angle of impact can have a significant effect on the answer and care needs to be taken in assessing this. The location and direction of the applied force both affect the precision of the calculation and the investigator needs to understand these effects and interpret the results accordingly. Although reasonably reliable results can be obtained for a collision involving two cars, big differ-

Figure 2 The relative position of this car and the one that struck it can be deduced from the damage to the front. © Metropolitan Police Service.

ences in the weights of the vehicles can cause unacceptably large errors particularly for the lighter one. This error is proportional to the mass ratio of the vehicles and if this exceeds 2:1 calculated values should be treated with circumspection.

Pedestrian impacts

Interpreting the results of impacts between pedestrians and cars is particularly difficult and precise estimates of impact speed cannot be made. A variety of techniques are available which provide some guide to the speed of a car which strikes a pedestrian. When the pedestrian is struck by the front of the vehicle empirical relationships between the distance the pedestrian was thrown and the impact speed have been deduced (**Fig. 3**). Secondly, there is also a statistical correlation between impact speed and the severity of the pedestrian's injuries. Although this cannot provide unequivocal information about the speed of the vehicle it provides a means of assessing the reliability of other estimates. Finally the nature and location of the damage to the vehicle also provides an indication of the impact speed. None of these methods on their own provide the level of certainty that can be obtained from impacts between two vehicles but in combination they can give a clear indication of likely speed and provide a useful check on other data sources.

Figure 3 Pedestrian throw distance.

Vehicle condition

Although a very high proportion of road traffic incidents are caused by road users, defective vehicles do contribute to these incidents. The defects can be caused by poor maintenance, prior damage or modifications which are ill conceived and/or poorly executed. An important part of any investigation is the elimination of potential causes and a careful examination of the vehicles involved for defects which might have contributed to the incident should always form a part of the inquiry. This examination is complicated by vehicle damage which coincides with the apparently defective component. This is not always a coincidence as more often than not the failure of the component is a consequence and not a cause of the incident.

Photographs and plans

Photographs and plans are an invaluable source for the investigator and one is not a substitute for the other. If the investigator did not visit the scene, photographs provide a source of information which supplements and amplifies that obtained from witnesses and provide an indispensable aide memoire in other cases. It is not unknown for tire marks which could not be seen clearly at the scene to show up in photographs. However, photographs can distort as well as enhance perceptions. Camera angle and the lens used may distort perspective and liberal use of a flash gun will make the scene appear brighter than it really was.

Plans are only as good as the survey from which they were drawn. However, they should show the relationship between the vehicles, victims, marks and ephemeral evidence and the important features of the scene accurately. They can be used to derive measurements which were not taken at the scene although all important measurements should have been recorded directly at the scene. The relationship between the scale of the plan and the error in such derived values needs to be borne in mind. Rectification programs are now available which allow information to be extracted from photographs and complete plans constructed or missing detail added to existing plans.

Accurate plans are an essential starting point for any simulation program.

Accurate records that can be understood by a layman will be important in any subsequent court proceedings. Accuracy will help to insure that records made by other personnel correspond with the plans.

Speed

Why calculate speed

Why does estimating vehicle speed figure so predominantly in road traffic incident investigation? There are two primary reasons for this. First, speed itself can cause the crash. A good example of this is loss of control when cornering. Such accidents are usually caused by traveling too fast for one or all of these reasons:

- the ability of the driver;
- the performance of the vehicle;
- the condition of the road.

This is most clearly seen and most easily dealt with when the car leaves the curved striated tire scuff marks characteristic of a vehicle cornering at the limits of adhesion.

The other principal reason for wanting an estimate of speed is to use it as a factor in the analysis of the incident. In any collision involving two moving objects the analysis needs to consider their relative movements and intervisibility which requires some knowledge of their velocity.

Vehicle speed

There are many ways in which the speed of a vehicle can be estimated. These include:

- the length of skidmarks;
- the radius of curved scuffmarks;
- vehicle damage.

Skidmarks Skidmarks are left by wheels which are no longer rotating. The marks are characteristic in appearance (**Fig. 4**) and caused as the wheels slide across the surface of the road. If all the wheels of the vehicle have locked then it is only friction between the tires and the road surface that is slowing the vehicle down. Although a great variety of tires are available those designed for road-going cars are all subject to the same design constraints. As a consequence, the major variables which determine the distance a car takes to skid to a halt are speed and the nature of the road surface. Thus if the coefficient of friction (μ) between tires and the road surface can be measured then a reliable estimate of the speed (v) can be calculated from the length of the tire marks (s).

Figure 4 This car left two clearly defined skid marks which deviate sharply to the left part way along. This indicates the point of impact with the victim. © Metropolitan Police Service

If a car is skidded to a halt from a known speed then:

$$\mu = v^2/2gs$$

where g is the acceleration due to gravity and then the speed of the car leaving the skidmarks is given by:

$$v = \sqrt{(2\mu gs)}$$

Curved scuffmarks The cornering force required to enable a car to follow a curved path is generated by friction between the tires and the road surface. Consequently tire/road friction sets an upper limit on the cornering force which can be generated and hence the speed at which any particular curved path can be followed. If this maximum speed is exceeded then the car side slips and leaves characteristic tire marks. These marks are curved and have a pattern of diagonal striations across them (**Fig. 5**). The rear wheel on the outside of the curve also tracks outside the front wheel. Once a car starts to leave these scuffmarks the driver has lost control.

Experimental evidence shows that once this point is reached the difference in handling between different vehicles has no effect on the radius of the first part of the scuffmarks which are left. These differences may, however, affect the driver's ability to regain control of his or her vehicle. The significant variables are the coefficient of friction and the radius of the tire marks.

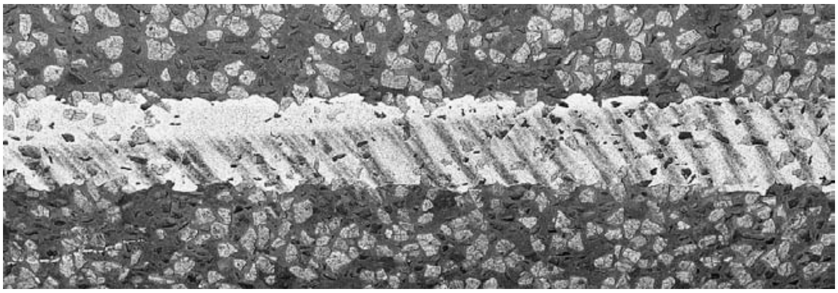


Figure 5 A vehicle cornering at the limits of adhesion leaves characteristic striated scuff marks which show clearly on the white line in this photograph. © Metropolitan Police Service.

If these are known, the speed of the car can be estimated with a precision of about $\pm 10\%$. The radius (r) of the initial section of the scuffmark can be obtained by simple geometric methods and then:

$$v = \sqrt{(\mu g r)}$$

Vehicle damage At a very simple level there is no surprise in the observation that the faster a car is traveling when it hits something the greater the damage. The observation that the extent of the damage to vehicles of similar size in similar impacts is comparable is perhaps rather more unexpected. The explanation is relatively simple, the manufacturers all have to face the same design constraints and type approval testing. This provides another useful tool for determining the speed of vehicles in road traffic incidents. Vehicle damage alone will not enable the impact speed of a car to be calculated as the extent of the damage depends on the change in speed (Δv) brought about by the collision. Thus a car brought to a halt by colliding with a wall at 30 mph will receive the same damage as a car colliding with another, stationary, car of similar mass at 60 mph where the post impact speed of both vehicles will be 30 mph. In all three cases Δv is 30 mph.

Although it is possible to calculate Δv by hand most of this work is done using computer programs. The data supplied to the program must be accurate and like all computer models the user must understand the limitations. In knowledgeable hands these programs can give accurate and reliable estimates of Δv which, coupled with knowledge of the post impact behavior of the car(s), can provide an estimated impact speed.

Pedestrian speed

If the conflict which led to the incident being investigated involved a pedestrian some estimate of the pedestrian's speed of movement will be required. This can be arrived at in a number of ways. An upper limit on how fast a pedestrian could possibly be moving can be deduced from the world record for the 100 m race. As this is just under 10 s no ordinary pedestrian is likely to run at more than $10\,\mathrm{m\,s^{-1}}$. Of more direct relevance the rule of thumb (Naismith's Rule) used by hill walkers to calculate point to point times is based on a walking speed of 3 miles or $5\,\mathrm{km\,h^{-1}}$ – equivalent to $1.4\,\mathrm{m\,s^{-1}}$ – on the level. This is probably a good starting point for any calculations based on a pedestrian's walking speed.

An alternative approach is to look up previously recorded figures for pedestrian speed. These may provide information about the affects of age, injury and disability on walking speed but some effort will be required to interpret the information given.

Finally the speed at which people similar to the victim move can be measured. This can be tackled in a number of ways. For example covert observation of the pedestrian crossing where the incident occurred will enable an investigator to measure the time taken by people similar to the victim to cross the road. An alternative, particularly where a child running was involved, is to get a number of children of the same age and stature as the victim to run over the same distance and to time them. This should be treated as a game if possible and is best done in consultation with the adult responsible for the children. Although asking the victim to repeat his or her movements has obvious appeal it should be remembered that the victim is not an independent observer of events.

Time and Distance

In many cases the reason for determining the speed of both vehicles and pedestrians is to analyze the

movements of the two objects which collided. Such analysis will show where the car might have been when the pedestrian stepped from the kerb or where the vehicle in the main road was when the other vehicle pulled out from the side turning. Comparison of the impact speed of an emerging vehicle with the distance traveled from a stop line to impact may enable the investigator to draw conclusions as to whether or not this vehicle complied with the stop sign. Two other factors, visibility and reaction time, play an important part in assessing the relative movements of vehicles and pedestrians. Without some knowledge of these it is impossible to draw conclusions from the relative movements of the parties to the incident.

Visibility

Before a driver or pedestrian can react to the presence of someone else they have to be able to see them. Although measuring visibility is not difficult it needs to be done with some thought. For example the visibility for one party may not be the same as that for the other. Take the example of a car waiting to emerge from a side turning where visibility is restricted by a fence. If the car has a long bonnet the driver will be some distance behind the give-way or stop line and thus the driver of the oncoming vehicle will be able to see the waiting car before the waiting driver can see the approaching vehicle (**Fig. 6**).

The height of the vehicle and the driving position and even the height of the pedestrian can have a significant effect on visibility. Thus when measuring visibility the observer should have his or her eyes at the same height as the witness and should be in the same lateral position in the road. As far as possible the target should be as similar as possible to the actual target in the incident.

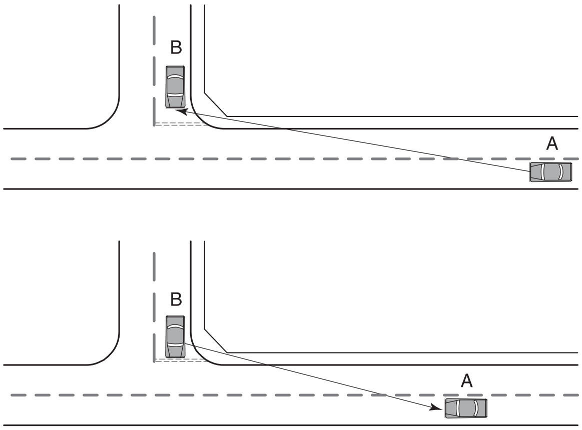


Figure 6 Visibility may not be the same for both vehicles involved in an incident. Top: car A is able to see car B from some distance before it reaches the 'Give Way' sign. Bottom: Even at the stop line car B cannot see car A until it is much closer.

Reaction time

A ball rolls into the road with a child in hot pursuit. This is a clearly identifiable hazard but there is an appreciable delay before the approaching car begins to brake. Why was there a delay between the hazard appearing and the driver's response? Human beings cannot react instantly because a number of cognitive processes have to take place first.

- Perception of the child approaching or entering the carriageway;
- Recognition that a hazard has appeared;
- Decision as to whether to brake or steer;
- Action to apply the brakes.

For most drivers the time taken to complete the last two stages will be very short. In the context of the investigation of dangerous occurrences on the road, however, it is the delay in seeing and recognizing the hazard which gives rise to arguments. This delay between the hazard appearing and action commencing is called reaction or perception–response time.

Although defining reaction time is a necessary preliminary to measuring it, actual measurement is difficult. Most systems designed to measure reaction time modify the value being measured by warning the driver to expect something and telling him or her what action to take. Many tests bear little resemblance to real driving conditions. The most widely known estimate of reaction time in the UK is that used to derive the thinking distance component of stopping distance in the UK's Highway Code. The origin of this figure, 0.68 s, is lost in the mists of time but it is probably significant that it amounts to 1 foot of thinking distance for every mile per hour of speed. It is unrealistically short for normal drivers in real driving conditions.

It is difficult to set up an experiment which measures driver reaction time without having to tell the driver what is happening. One attempt to do so used a driving simulator for what were ostensibly tests quite unconnected to reaction time. Eleven drivers were tested who fell into two groups with mean reaction times of 0.83 s and 1.13 s. The slowest driver took 1.43 s. Alcohol and some other drugs tend to extend a driver's reaction time, often quite significantly. Perception–response time is discussed at some length in the literature.

Although driver reaction time is a complex problem in cognitive psychology, investigation of road traffic incidents requires an understanding of the concept and the range of values which the normally

alert driver might achieve. The effect of expectation on response should not be overlooked. Drivers do not expect pedestrians to be crossing multi-lane dual carriageways and thus react slowly when they do.

Interpretation

General considerations

The critical part of any investigation is the collation and interpretation of the information obtained. This should be tackled in a logical and methodical manner and will be an iterative process which should have started as the information is collected. A growing understanding of the incident will have directed the investigator's inquiry to insure that the right information is obtained and nothing is overlooked. Care must be taken not to exclude alternative plausible explanations until sufficient information has been collected to justify ignoring them.

All evidence particularly that of eye witnesses should be considered critically.

- Could the witness see?
- What attracted the witness's attention?
- Are the plans accurate?
- Do the photographs give the correct impression with no distortion of perspective.
- Is what the witness saw possible?
- What is the precision of the calculations and how in consequence should they be interpreted?

Analysis of the information will go through a number of overlapping stages. Information obtained from one stage may well modify that obtained from an earlier stage which will need to be re-assessed. By the end of this iterative process the investigator will have firm factual and logical foundation for his or her conclusions.

Analytical process

There is no one correct way of analyzing an incident but all of the following steps should be included.

Plans, photographs and observations Good plans and photographs enable the investigator to get a clear picture of the scene, an essential first step in the analysis of the incident. Plans and photographs:

- provide a vital aid to understanding what the witnesses are saying;
- show the relationship between marks, participants and the environment;
- allow the investigator to develop his her own interpretation of events;
- may be used as a source of data.

The eye witnesses The evidence of eye witnesses needs to reviewed carefully. However, with care they can provide reliable evidence which can be obtained in no other way. The investigator needs to consider the following.

- What do they say and are they consistent?
- If, as is normal, they disagree are there any areas of agreement?
- Are there obvious explanations for disagreements (e.g. their viewpoint)?
- Is there information which will be required for subsequent analysis? e.g. the state of traffic signals; the direction of travel and speed of pedestrians and vehicles.

Skidmarks and scuffmarks Ephemeral evidence left at the scene is one of the basic building blocks of a successful reconstruction of a road traffic incident.

- What marks have been left on the road surface?
- What can be deduced about the movements of vehicles and pedestrians from them?
- Can they be used to estimate speed?

The vehicle(s) The condition of the vehicle(s) involved in the incident is important for two reasons.

- Is the vehicle itself a cause of the incident?
- Can anything be deduced about impact speed?

Calculations Calculations based on observations and measurements are inevitably multistage processes which may involve re-evaluation of the other information between stages:

- Estimate of vehicle speed from tire marks, damage etc.
- Estimate time taken for:
  - the pedestrian to cross the road;
  - another vehicle to emerge from a side road;
  - a vehicle to skid to impact.
- Investigate questions of intervisibility or reaction time.

With all calculations the investigator needs to be aware of:

- the accuracy and precision of the result;
- the sensitivity of the result to changes in input information.

At each stage other information should be re-evaluated and the direction of the investigation reviewed. What should be done and in what order will depend on the circumstances, but at the end of one or more iterations the investigator should have a clear understanding of the incident and how it

developed and be in a position to explain it in clear, simple language.

Conclusion

Throughout any analysis of an incident the investigator should always search for ways of corroborating one source of information with another. At the end of the process the explanation of the sequence of events leading up to and through the incident should be internally consistent and include all the known data. If some data have been excluded the investigator should provide reasons – no data should just be ignored. Finally, the investigator must report his or her findings clearly and simply setting out the information available, logical process gone through, and the conclusions reached. Just as photographs, plans and other illustrations can assist the investigator to understand what happened they can be used to support and amplify the report. The old adage that a picture is worth a thousand words remains true.

The primary reason for investigating road traffic incidents is to learn what lessons they teach and thereby improve road safety. The investigator endeavors to establish what happened, why it happened and how repetition can be avoided. By learning from experience the UK has reduced annual road traffic fatalities from around 5600 to around 3600 in 13 years. Further significant reduction in this figure must be achieved and thorough, knowledgeable investigation of all incidents and application of the lessons learnt will continue to play a significant part in achieving this goal.

See also: **Accident Investigation**: Motor Vehicle; Rail; Determination of Cause: Reconstruction; Tachographs; Driver Versus Passenger in Motor Vehicle Collisions. **Pattern Evidence**: Vehicle Tire Marks and Tire Track Measurement.

Further Reading

Allen M, Abrams B, Ginsberg A and Weintrub L (1996) *Forensic Aspects of Vision and Highway Safety*. Tucson, AZ: Lawyer's & Judge's Publishing.

Backaitis SH (1989) *Reconstruction of Motor Vehicle Accidents – A Technical Compendium*. Warrendale, PA: Society of Automotive Engineers.

Backaitis SH (1990) *Accident Reconstruction Technologies – Pedestrians and Motorcycles in Automotive Collisions*. Warrendale, PA: Society of Automotive Engineers.

Baker JS and Fricke LB (1986) *Traffic Accident Investigation Manual – At Scene and Technical Follow Up*. Evanston, IL: Traffic Institute, Northwestern University.

Eubanks JJ (1994) *Pedestrian Accident Reconstruction*. Tucson, AZ: Lawyer's & Judge's Publishing.

Fricke LB (1990) *Traffic Accident Investigation Manual*; vol. 2, *Traffic Accident Reconstruction*. Evanston, IL: Traffic Institute, Northwestern University.

Noon RK (1994) *Engineering Analysis of Vehicular Accidents*. Boca Raton, FL: CRC Press.

Obenski KS (1994) *Motorcycle Accident Reconstruction – Understanding Motorcycles*. Tucson, AZ: Lawyer's & Judge's Publishing.

Olsen PL (1996) *Forensic Aspects of Driver Perception and Response*. Tucson, AZ: Lawyer's & Judge's Publishing.

Watts AJ, Atkinson DR and Hennessy CJ (1996) *Low Speed Automobile Accidents*. Tucson, AZ: Lawyer's & Judge's Publishing.

Determination of Cause: Reconstruction

H Steffan, Technical University Graz, Graz, Austria

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0395

Introduction

Car traffic has increased dramatically. Although additional traffic regulations, such as speed limits, and other restrictions were imposed to reduce accidents, crashes can be seen every day. Many result in vehicle damage or injuries to occupants or others. Claims resulting from these accidents often have to be settled in court.

In most countries special 'accident experts' assist judges in reconstructing accident situations, based, for example, on information from eyewitnesses, car deformations or tire marks. In many countries these experts are either specially trained police officers or members of the road authorities. In others, reconstructions are performed by independent specialists. These accident analyses, which form the basis for court decisions, were the start of specific accident research.

Accidents are also reconstructed for many other reasons. Car manufacturers perform in-depth studies of real accidents to learn more about accident mechanisms and to study the potential and the effectiveness of new safety devices in cars. Accidents are also often reconstructed and analyzed to help improve road planning.

There are several questions that must be answered when reconstructing a car accident. Today's accident reconstruction tries to provide full information about the movement of all involved vehicles, persons or objects from the point of first visual contact to their

rest positions. Time history of velocities, positions and crash-related data, such as velocity changes, deformation energies and passenger loads, must be analyzed. In addition, prevention analyses are often included. These analyses allow determination of the conditions necessary to prevent repetition of the accident.

There is one general rule in any accident reconstruction: a detailed collection of scene data forms the basis for a good reconstruction. Some of the most important scene data are:

- vehicle or human rest positions;
- road marks;
- vehicle damages and marks;
- personal injuries.

Owing to the increased number of cars with ABS, fewer tire marks, which formed the most objective basis, can be seen and thus one major contribution to the reconstruction is lost. On the other hand, the increased performance of personal computers has made it possible to use highly sophisticated reconstruction or simulation algorithms to study accidents.

In recent years several software products have been developed for the reconstruction of vehicle accidents. Some of them are designed just to calculate vehicle velocities from the kinematics of specific accident types. Others, with full vehicle simulators, allow the simulation of car motion during the accident, starting from the point of reaction to the end position for any kind of accident. It is now even possible to visualize the car motion on the screen during the calculation in real time. Three-dimensional geometric information regarding road marks can be imported from CAD systems or from photographs as scanned bitmaps, and all important road, car and tire characteristics can be taken into account, as well as the reactions of the driver.

Targets

It is important, when formulating the physical and mathematical model of a car, to take into account the fact that many parameters are not well known when reconstructing the accident. In particular, the condition of the tires, such as their age or air pressure, the dampers or the steering system are often not well documented. The resistance of the chassis is only known for well-defined crash conditions and not for the specific accident situation. Therefore it is important to find a compromise between accuracy and the amount of input data necessary to perform the calculation. On the other hand, the user should be able to take into account all parameters which are documented and known to influence driving and crash behavior. Based on this knowledge, simplified models have to be used, which guarantee that the basic driving behavior is predicted correctly. All necessary input parameters must be defined on the basis of physical parameters and a simple means of varying the parameters guarantees that the expert can cross-check their influence on the simulation result.

The automotive industry uses several simulation models to learn about the driving behavior of vehicles. These industrial vehicle dynamics simulation programs are optimized to predict driving behavior under well-defined initial and boundary conditions. As a result these models require many input parameters. For the reconstruction of a vehicle accident, such detailed knowledge, especially regarding the suspension, the tires and the road conditions, is normally not available. In addition, the steering and degree of braking are often not known. It is thus difficult to use these simulation models for the reconstruction of vehicle accidents.

A similar problem exists regarding collision models. Several programs (mainly finite element (FE)-based) exist that allow the calculation of the deformations for well-defined collision conditions. To achieve good agreement with real impacts, they require a very detailed knowledge of the vehicle structure and a very powerful computer. Some 100 000 degrees of freedom are required to model one vehicle properly. In addition, every simulation has to be calibrated with similar crash tests. This is the major reason why FE programs are only used for accident reconstruction in a few specific cases.

Several computer programs have been developed especially for the reconstruction of vehicle accidents. They allow the calculation of vehicle motion and collisions based on various physical models. PC-CRASH is one of these programs with a worldwide distribution. It uses a kinetic time forward simulation of vehicle dynamics and combines it with a momentum-based collision model; accidents can be reconstructed, starting from the point of reaction to the end position, for all involved cars simultaneously. The reconstruction is performed in an interactive graphical environment, which allows a sketch or photograph of the accident scene to underlay the reconstruction. For an effective presentation of the results, 3D animations can be created directly from the calculated results.

Accident Analysis

Postcrash movement

In a conventional accident analysis the reconstruction is normally started from the rest positions of the involved vehicles and calculated backwards to the

collision position. To determine the postcrash velocity of the vehicles, an average deceleration has to be estimated. In many cases, where no tire marks are available, straight-line movement of the vehicle is assumed for determining the distance of postcrash travel. Depending on the road situation and involved vehicle, an average deceleration level is assumed, typically in the range of 0.1–10 m s^{-2}. The postcrash velocity can than be calculated according to the formula:

$$\nu = \sqrt{2as + \nu_0^2} \qquad \textbf{(Equation 1)}$$

where v represents the postcrash velocity (m s^{-1}), a the average deceleration (m s^{-2}) and s the postcrash travel of the vehicle's center of gravity (m).

The major problem with this method is estimating the vehicle's average deceleration during a complicated postcrash movement. In this phase the vehicle may have been rolling or sliding and this has a huge influence on the amount of energy dissipated. To overcome this problem, vehicle simulators are often used to compare the vehicle's movement with marks found on the road or on the vehicle. Thus the postcrash velocity can be determined more accurately.

Collision model

During the collision the contact forces between the two involved vehicles vary over time. These forces depend on the vehicle structure, the deformation velocity, the contact situation and several other parameters. As these dependencies are highly nonlinear and very difficult to formulate, the treatment of the parameters through their integral values has proven to be more efficient. Modern FE programs allow investigation of the time variation of the contact forces. But these programs require enormous calculation times and a huge amount of modeling. Simplified models like the CRASH algorithm have produced large errors under several conditions. In many cases, insufficient knowledge regarding the structural deformation behavior is available to estimate the proper parameters.

Crash hypotheses that only compare the velocity conditions of both vehicles before and after the impact have therefore been used with great success for the reconstruction of accidents.

Material properties in the contact area During the contact phase large forces may occur, which cause deformations to one or both collision partners. These deformations may remain fully or partially after the impact (plastic deformations) or they may fully recover (elastic deformations). Through the definition of the parameter k the amount of elasticity for a crash can be defined. This parameter k is only valid for a whole crash situation and not for one of the partners. Therefore one crash partner may exhibit a high degree of elasticity when impacting with partner A and a high degree of plasticity when impacting with partner B.

Using an example of a straight central impact (**Fig. 1**), the parameter k can be easily explained. To insure that the two partners collide, the velocity of partner 1 must be higher than the velocity of partner 2. In phase 1 contact forces act, which reduce the velocity of mass 1 and increase that of mass 2. They are equivalent in size, but with opposite direction. The relation between acceleration and contact force is calculated from:

$$ma = F \qquad \textbf{(Equation 2)}$$

where m defines the mass, a the acceleration and F the contact force. The end of phase 1 is defined when both partners have equal velocities. It is called the 'compression' phase. In phase 2 the forces reduce again until the two masses separate. This phase 2 is called 'restitution'.

In this case the coefficient of restitution is defined as

$$\underline{k} = \frac{\nu_2' - \nu_1'}{\nu_2 - \nu_1} \qquad \textbf{(Equation 3)}$$

$0 < k < 1$ where k=0 represents a fully plastic and k=1 a fully elastic impact.

Eccentric impacts **Figure 2** shows an example of a measurement of the two acceleration components for an eccentric impact. The compression moment S_c and restitution momentum S_R are now defined as

$$S_C = \int_{t_0}^{t_m} Fdt \qquad \textbf{(Equation 4)}$$

$$S_R = \int_{t_m}^{t_1} Fdt \qquad \textbf{(Equation 5)}$$

where t_0 is the time of first contact, t_m defines the time of identical velocity for both vehicles at the point of impact and t_1 describes the time of separation.

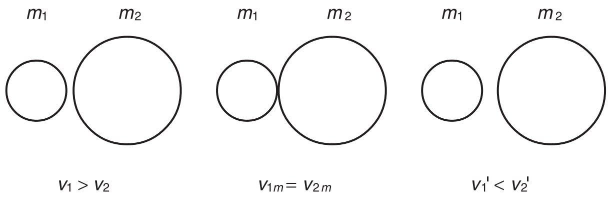


Figure 1 Central impact. See text for explanation.

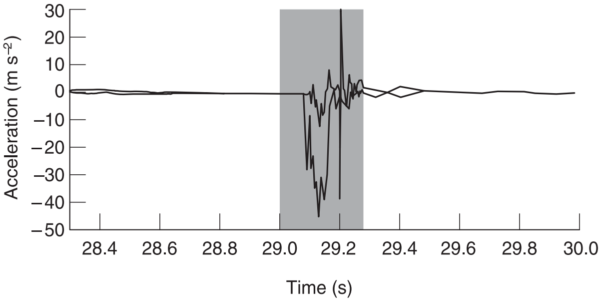


Figure 2 Measured accelerations.

In the case of a generic eccentric impact, the 'coefficient of restitution' is now defined as a ratio between restitution momentum and compression momentum

$$\varepsilon = \frac{S_R}{S_C} \qquad \textbf{(Equation 6)}$$

The total exchanged momentum is calculated from

$$S = S_C + S_R = S_C(1 + \varepsilon)$$

In a full impact the velocities of both vehicles at the impulse point must be identical at the end of the compression phase (**Fig. 3**). For simplicity, the impact model is only derived here in 2D. In some programs like PC-CRASH the models are also extended to 3D. They are identical except that all velocities, the momentum and angular momentum are defined in 3D coordinates and all three components of the velocity and rotational velocity are taken into account.

As can be seen in Fig. 3 a local coordinate system is defined, which originates at the 'impulse point'. The components of the relative velocity for both vehicles at the impulse point can be calculated from:

$$V_{1t} = v_{s1t} + \omega_{1z} n_1 \qquad \textbf{(Equation 8)}$$

$$V_{1n} = v_{s1n} + \omega_{1z} t_1 \qquad \textbf{(Equation 9)}$$

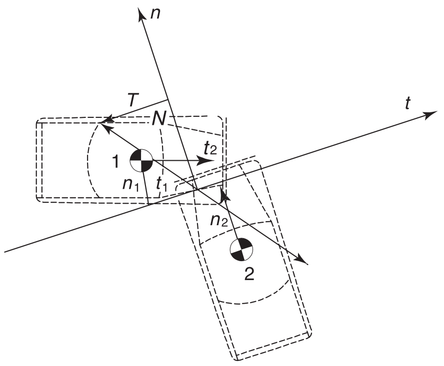


Figure 3 Impact configuration.

where V_{lt} defines the velocity component of the impulse point for vehicle 1 in direction t and V_{ln} in direction n. So the components of the relative velocity for both vehicles at the impulse point can be calculated from

$$V_t = V_{1t} - V_{2t} \qquad \textbf{(Equation 10)}$$

$$V_n = V_{1n} - V_{2n} \qquad \textbf{(Equation 11)}$$

In addition, the balance of momentum can be formulated for both vehicles

$$m_1(v'_{s1t} - v_{s1t}) = T \qquad \textbf{(Equation 12)}$$

$$m_1(v'_{s1n} - v_{s1n}) = N \qquad \textbf{(Equation 13)}$$

$$m_2(v'_{s2t} - v_{s2t}) = -T \qquad \textbf{(Equation 14)}$$

$$m_2(v'_{s2n} - v_{s2n}) = -N \qquad \textbf{(Equation 15)}$$

The balance of angular momentum can be formulated

$$I_{1z}(\omega'_{1z} - \omega_{1z}) = Tn_1 - Nt_1 \qquad \textbf{(Equation 16)}$$

$$I_{2z}(\omega'_{2z} - \omega_{2z}) = -Tn_2 + Nt_2 \qquad \textbf{(Equation 17)}$$

When combining these equations, the change of the relative velocity for both vehicles at the impulse point can be calculated from

$$V'_t = V_t + c_1 T - c_3 N \qquad \textbf{(Equation 18)}$$

$$V'_n = V_n - c_3 T + c_2 N \qquad \textbf{(Equation 19)}$$

with

$$c_1 = \frac{1}{m_1} + \frac{1}{m_2} + \frac{n_1^2}{I_{1z}} + \frac{n_2^2}{I_{2z}} \qquad \textbf{(Equation 20)}$$

$$c_2 = \frac{1}{m_1} + \frac{1}{m_2} + \frac{t_1^2}{I_{1z}} + \frac{t_2^2}{I_{2z}} \qquad \textbf{(Equation 21)}$$

$$_3 = \frac{t_1 n_1}{I_{1z}} + \frac{t_2 n_2}{I_{2z}} \qquad \textbf{(Equation 22)}$$

To be able to solve these equations and to calculate the postimpact velocities and rotations two additional assumptions have to be made (see below).

The full impact In case of a full impact two additional assumptions are made:

1. No relative movement between both vehicles can be found in the impulse point at the end of the compression phase.

$$T_c = \frac{V_n c_3 + V_t c_2}{c_3^2 - c_1 c_2} \qquad \textbf{(Equation 23)}$$

$$N_c = \frac{V_n c_1 + V_t c_3}{c_3^2 - c_1 c_2} \qquad \textbf{(Equation 24)}$$

2. The average between compression and restitution momentum is defined by the coefficient of restitution, which is defined according to **Equation 6**.

So the components of the total momentum can be calculated from

$$T = T_c(1+\varepsilon) \quad \textbf{(Equation 25)}$$

$$N = N_c(1+\varepsilon) \quad \textbf{(Equation 26)}$$

These equations are sufficient to calculate all post-impact velocity conditions for both involved vehicles in the case of a full impact.

The sliding impact In certain collisions the two vehicles will never reach identical velocities at the impulse point during the impact. In such cases a contact plane has to be defined, along which the two vehicles slide. The impulse point must coincide with this plane. For such a situation the following two assumptions are made:

1. No relative movement between both vehicles can be found at the impulse point at the end of the compression phase in a direction normal to the contact plane. So N_c can be calculated from **Equation 24**.
2. The direction of the momentum is limited by a friction (μ). This value defines the friction between the two impacting vehicles.

$$T = \mu N \quad \textbf{(Equation 27)}$$

3. The average between compression and restitution momentum is again defined by the coefficient of restitution according to Equation 6, and T and N can again be calculated from **Equations 25** and **26**.

The coefficient of restitution, which can be handled as an input or output parameter, is easy to define. It always lies in the range 0.1–0.3. The higher the deformations of the vehicles, the lower is the coefficient of restitution. Only for very small deformations are values higher than 0.3 possible.

Using these equations, the relation between post-crash and precrash velocities can be determined. These equations are used in many different forms to calculate the precrash velocities from the postcrash velocities, mainly for the conventional reconstruction of car accidents, where the reconstruction is started from the rest position. Several simulation models use the same equations to determine the postcrash velocities from the precrash values.

Energy equivalent speed As modern cars are mostly designed to absorb energy during a collision, the amount of damage found on the vehicles can also be used to determine the precrash velocity. For most vehicles crash tests are performed and the results published. They show the vehicles' deformation after an impact with a well-defined barrier. **Figure 4** shows the deformations of a certain car type in the frontal area when impacting a rigid barrier with speeds of 15 and 48 $km\,h^{-1}$. Significant differences can be found in the amount of damage sustained by the vehicle. This knowledge can also be used for accident reconstruction.

When comparing the deformations found on the vehicle to those of the reference tests, the amount of energy absorbed by the vehicle due to the impact can be estimated. As deformation energies are not very convenient quantities, they are recalculated into velocities and called energy equivalent speed (EES). EES is defined thus:

$$EES = \sqrt{\frac{2E_{Def}}{m}} \quad \textbf{(Equation 28)}$$

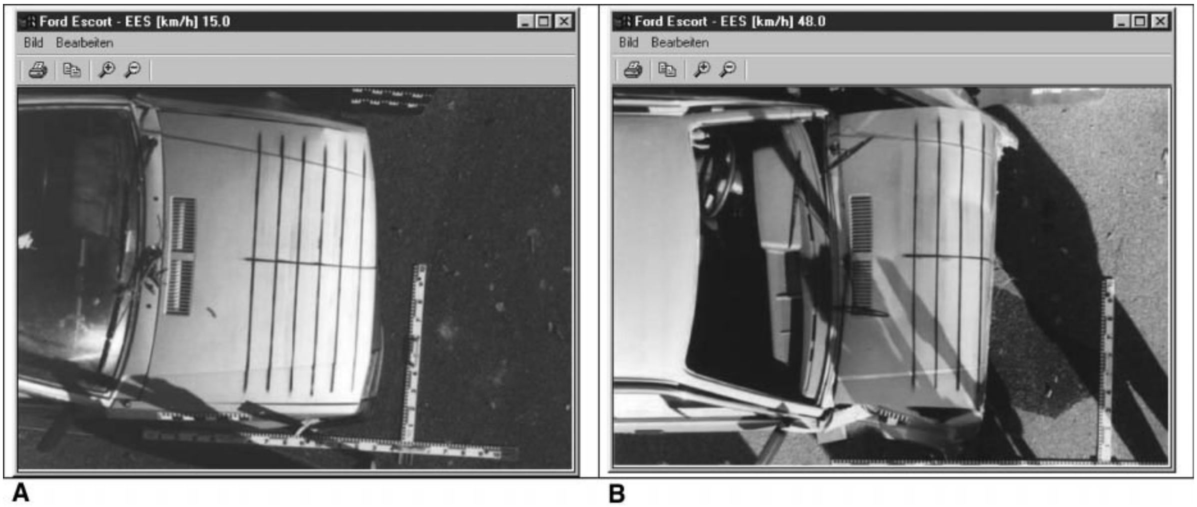


Figure 4 Vehicle damage after frontal impact at (**A**) 15 $km\,h^{-1}$, and (**B**) 48 $km\,h^{-1}$.

Where E_{Def} defines the vehicle's deformation energy and m the vehicle's mass.

Thus the conservation of energy can be used in addition:

$$E_{\mathrm{kin}1}{}' + E_{\mathrm{kin}2}{}' = E_{kin1} + E_{kin2} - E_{Def1} - E_{Def2} \qquad \textbf{(Equation 29)}$$

where E_{kini} represents the kinetic energy of vehicle i before and E_{kini}' after the impact.

Precrash movement

As in many cases the impact speed is not the relevant speed, precrash analyses are also of great importance. They allow determination of the drivers' or other involved persons' reactions. The velocities, relative positions and visibilities at the reaction point can then be investigated and the cause of the accident and the failures of the involved persons or vehicles can be identified. There are several types of failures or combinations of failures, that cause accidents: the vehicle velocity not being adapted to traffic conditions and insufficient driver attention are only just two examples. Whenever accidents are investigated, the situation has to be analyzed from the reaction point. Only thus can real cause be found.

The Driver's Reaction

The point of reaction can be determined by various methods, depending on the accident situation. In principle, the time from the reaction to the collision position normally consist of three major phases:

- reaction time;
- lag;
- action time.

Reaction time

The reaction time defines the necessary time to identify the situation, decide the kind of reaction and start the action through activation of certain muscles. It depends on many factors, like age and tiredness, and also on some which are difficult to estimate; for example, an eye blink may increase the reaction time by approximately 0.2 s. One aspect, which is also of great importance, is the visibility of the object. In cases of low contrast or small light differences, reaction time may increase dramatically. This is why so many pedestrian accidents occur at night. In addition, there is a significant difference if the resulting action has to be performed by the arms or by the legs. The greater distance between brain and leg means that an action being performed with the leg will require a significantly longer reaction time.

Typical reaction times are between 0.5 and 1.5 s. Racing drivers are known to react during a race within a time of 0.3 s.

Lag

The lag is defined by the amount of time required by the technical system to act. There is no lag for the steering system but a lag of 0.1–0.3 s for the brake system. This lag is mainly caused by the amount of time necessary to push the brake pedal to such a level that the full brake force is applied to the wheels. One technical solution to reducing this time is the so-called 'brake assistant', which automatically applies the full brake force when the brake pedal is pushed very fast.

Action time

The action time defines the time when the action (mainly braking or steering) is active.

Prevention analysis

When the point of reaction has been found, so called prevention analysis must be performed. These calculations are used to determine a fictitious scenario where the accident would not have occurred. In this case, parameters like initial velocity, shorter reaction time, better visibility or higher friction on the road are varied. These analyses are used both to determine the influence of the driver's failure and often to change the road design to prevent similar accidents.

Sample Case

The following sample case demonstrates the analysis of a vehicle–vehicle accident caused mainly by one driver overrunning a 'stop' sign.

The first step when simulating a new accident is the identification of all involved vehicles. Modern simulation programs to allow access various databases containing all necessary geometric and mass data.

In a second step the involved cars can be moved to the 'average' collision position. Here their correct overlap must be taken into account.

To define the friction conditions and drivers' actions, so-called 'sequences' can be given (**Fig. 5**). They can be defined in a very simple and flexible way. The different steering, brake and acceleration actions can be defined by listing the actions. The values for the individual actions can then be given. The validity of one sequence can be limited by definition of a time interval or a travel distance of the vehicle's center of gravity. The brake or acceleration forces can be given for every wheel independently. This is important to simulate wheels locked during the accident. Changes in the friction due to oil or wet areas can be defined by identifying individual areas. The corresponding friction coefficient must then be specified.

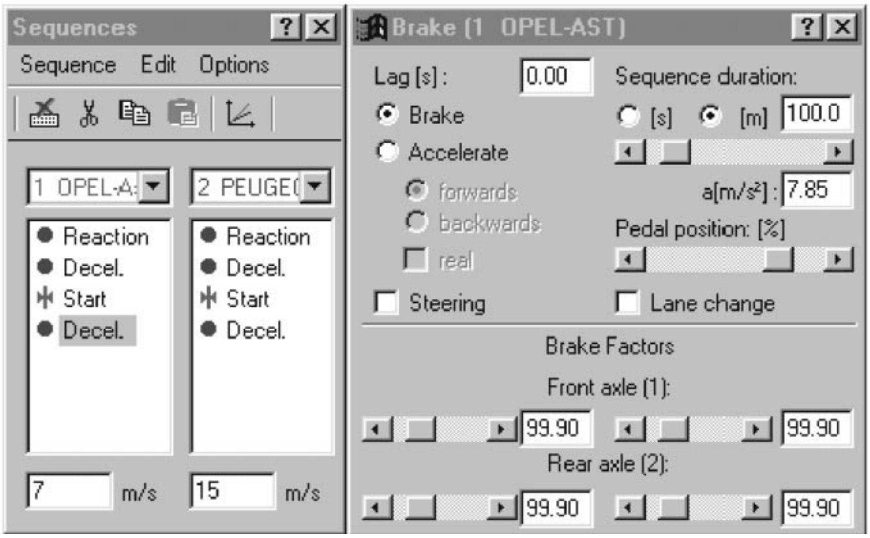


Figure 5 Definition of driver actions.

After these definitions have been given, the impact can be calculated and the postcrash movement will be simulated automatically (**Fig. 6**). As a result the movement of the involved vehicles, including their wheel traces, can be seen on the screen and can then be compared to a previously created DXF drawing of the scenario. As an alternative a scanned bitmap can be underlayed. So the preimpact velocities can be varied, as well as all other impact parameters. PC-CRASH, as one of the major software tools, also provides an optimization tool that allows automatic variation of specific parameters. Through the definition of target functions, the most plausible solution will be found immediately.

Figure 7 shows the movement of two vehicles after a 90° impact in steps of 200 ms. A dry asphalt surfaced road was assumed.

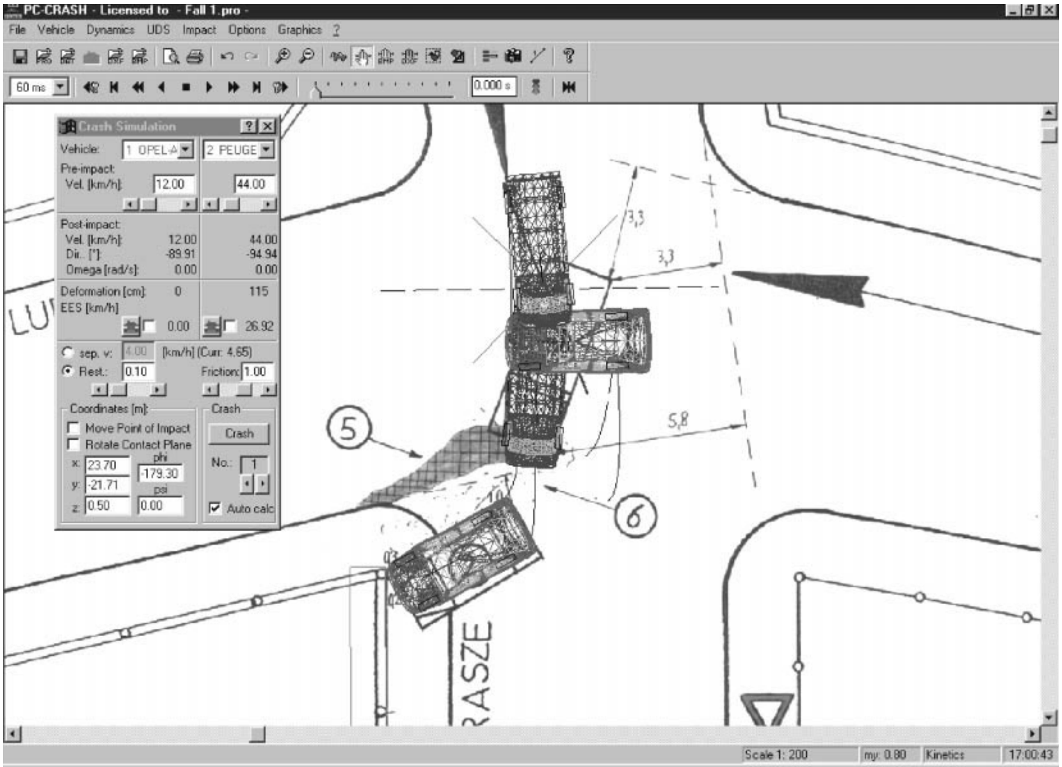


Figure 6 Definition of impact values including results.

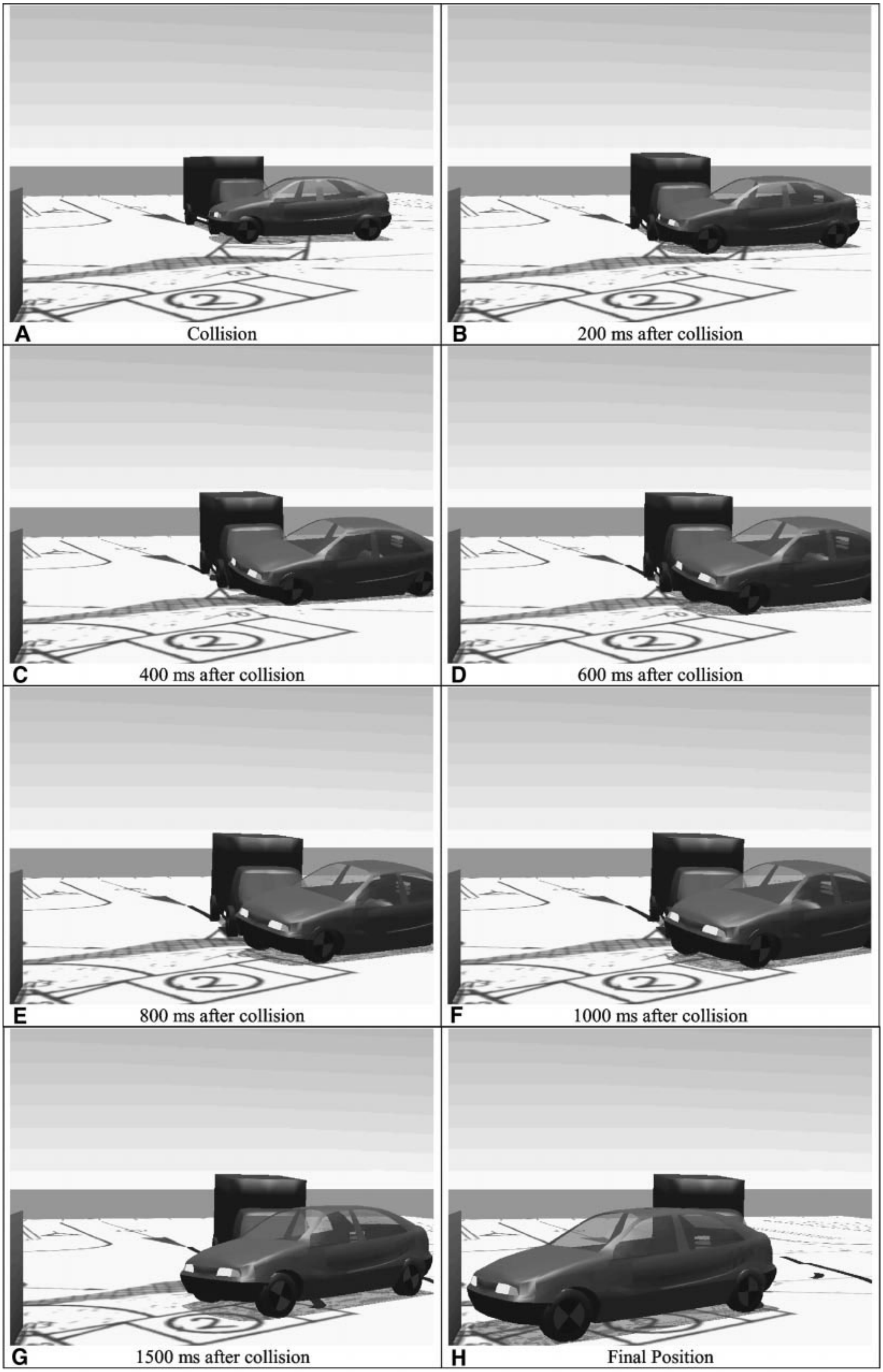


Figure 7 Vehicle movement after impact: (**A**) collision, (**B**) 200 ms, (**C**) 400 ms, (**D**) 600 ms, (**E**) 800 ms, (**F**) 1000 ms, and (**G**) 1500 ms after collision, (**H**) final position.

Results

The major fault in this case was that the driver of the private car should have stopped before the crossing. From brake traces drawn from the truck before the impact, which had a length of 7 m, the initial speed of the truck was calculated to be 60 km h^{-1}. The truck driver reacted 1.7 s before the collision. Driving at a speed of 48 km h^{-1} or less, the accident would have been prevented, as the truck would have stopped before the impact. For these calculations a reaction time of 1 s and a brake lag of 0.2 s was assumed. If the driver of the truck had been driving at a speed of 50 km h^{-1}, the truck would have reached the collision position 2.4 s after he had reacted. This means the private car would have continued for 0.7 s and thus have passed an additional distance of approximately 2.3 m, insufficient to prevent the accident. So the accident would have also occurred at an initial speed of 50 km h^{-1} which was the speed limit at the reaction point but the impact speed would have been reduced from 44 km h^{-1} to approximately 15 km h^{-1}.

See also: **Accident Investigation**: Motor Vehicle; Determination of Cause: Overview; Tachographs; Driver Versus Passenger in Motor Vehicle Collisions. **Pattern Evidence**: Vehicle Tire Marks and Tire Track Measurement. **Photography and Digital Imaging**: Digital Imaging Enhancement.

Further Reading

Brach RM (1991) *Mechanical Impact Dynamics: Rigid Body Collisions*. Chichester: Wiley.

Day TD and Hargens RL *An overview of the way EDCRASH computes Delta-V*, SAE 820045. Warrendale, PA: Society of Automotive Engineers.

Gnadler R (1971) *Das Fahrverhalten von Kraftfahrzeugen bei instationärer Kurvenfahrt mit verschiedener Anordnung der Hauptträgheitsachsen und der Rollachse*. Thesis, University of Karlsruhe.

Kudlich H (1966) *Beitrag zur Mechanik des Kraftfahreug-Verkehrsunfalls*. Thesis, University of Vienna.

Lugner P, Lorenz R and Schindler E (1984) The connexion of theoretical simulation and experiments in passenger car dynamic. In: Hedrick HJ (ed.) *The Dynamics of Vehicles on Roads and on Tracks*, Lisse: Swets-Zeitlinger.

McHenry RR (1976) *Extensions and Refinements of the CRASH Computer Program. Part I: Analytical Reconstruction of Highway Accidents*. DOT-HS-5-01124.

McHenry RR, Segal DJ and DeLeys NJ (1967) Computer simulation of single vehicle accidents. *Stapp Car Crash Conference*, Anaheim, California, October 1967.

Rill G (1994) *Simulation von Kraftfahrzeugen*. Vieweg.

Slibar A (1966) Die mechanischen Grundsätze des Stoßvorganges freier and geführter Körper und ihre Anwendung auf den Stoßvorgang von Fahrzeugen. *Archiv für Unfallforschung* **2**.

Steffan H (1993) PC-CRASH, A Simulation Program for Car Accidents. *ISATA; 26th International Symposium on Automotive Technology and Automation*, Aachen.

Steffan H *PC-CRASH 6.0. A Simulation Program for Car Accidents: Users and Technical Manual*. Linz: DSD.

Steffan H and Moser A *The collision and trajectory models of PC-Crash*, SAE 960886. Detroit: Society of Automotive Engineers.

Steffan H and Moser A *The trailer simulation model of PC-Crash*, SAE 980372. Detroit: Society of Automotive Engineers.

Driver Versus Passenger in Motor Vehicle Collisions

W S Smock, University of Louisville, Louisville, KY, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0749

Introduction

The determination of a motor vehicle occupant's role in a vehicular collision is always an important medicolegal question. If the occupants of the vehicle have been ejected or relocated within the vehicle as a result of the vehicular dynamics of the collision, determining the occupants' role at the time of the collision may be difficult. The investigation that coordinates an examination of injury mechanisms, occupant kinematics, vehicle dynamics and the evaluation of trace evidence will facilitate the determination of an occupant's role. Such a determination is critical when criminal charges are considered. It is imperative that, when charging an individual, the investigating agency performs a thorough investigation in order to protect the innocent passenger from being falsely prosecuted as the driver.

Variables that Confound Occupant Role Determination

'I wasn't the driver' is a statement often made by occupants of an automobile which has been involved in a serious or fatal collision. It is also a statement that may or may not be true. When a surviving occupant presents to an emergency department, it is imperative that care be taken to ensure that any short-lived evidence found on the individual is recognized and

collected. It is also imperative that all of a patient's injuries be documented before their appearance is altered in the delivery of patient care. By preserving evidence and painstakingly documenting injuries, the investigating agency insures a high standard of investigation. Statements obtained in the emergency department from the patients should be made part of the permanent record, within quotation marks, by those witnessing their utterances.

Allegations of criminal and/or civil liability should be based on the tangible evidence collected from the vehicles, their occupants and the crash scene. Common problems which hinder the investigative process itself often complicate the differentiation of driver from passenger. These impediments may also be used in court as an effective defense against the charges. The following are examples of potential problems:

- Occupants are removed from vehicles by well-meaning bystanders whose statements and observations are not recorded by the investigating agencies.
- Physical evidence of occupant movement within the vehicle and trace evidence of occupant and vehicle contact are not recognized, collected and documented.
- The surviving occupant's injuries, including pattern injuries, are not photographically documented while the person is in the emergency department.
- Evidence standards (clothing, shoes) and biological standards (hair and blood) are not collected from all occupants.
- An autopsy is not conducted on the deceased occupant(s).
- Substandard analysis of the crash scene prohibits accurate reconstruction of the vehicle dynamics.
- Vehicle components found at the crash scene are haphazardly thrown back into the vehicles, resulting in the possibility or supposition of cross-contamination.
- Inadequate resources are committed to the evaluation of the incident.
- Assumption is made that the owner of vehicle is the driver.
- The vehicle is not preserved or is left exposed to the elements, which may lead to destruction or loss of trace evidence.

Vehicle Dynamics

How the vehicle interacts with the environment will dictate what forces are applied to the vehicle's occupants. The forces applied to the occupants will dictate how the occupants move within the vehicle. When, for example, a vehicle hits a tree head-on, the occupants of the vehicle will move forward, toward the point of impact. This determination is best made by an accident reconstruction expert.

Occupant Kinematics

In the event of a rapid deceleration of the kind that occurs instantly in crash circumstances, vehicle occupants, restrained or unrestrained, will experience a force which causes them to move initially toward the primary area of impact. This occupant movement or occupant kinematics has been conceptualized as a motion parallel to and opposite from the direction of the force which is developed by the impacting object (**Fig. 1**). This movement of the occupant, and subsequent contact with the vehicle's components, is dictated by the forces applied to the vehicle through its interaction with the environment. The application of the principles of occupant kinematics will predict in what direction a particular occupant will move, and, therefore, which component within the vehicle they will strike. Occupant kinematics is best determined by a physician/nurse with forensic training in concert with the accident reconstructionist.

Injury Mechanisms

The correlation and matching of pattern injuries from surfaces and components within a vehicle will reveal an occupant's position during a portion of the vehicle's collision sequence. The most commonly observed pattern injuries are: seat belt contusions (**Fig. 2**); air bag injuries – facial abrasions, face and arm lacerations and fractures, and extremity amputations (**Figs 3–6**); steering wheel contusions; contusions, abrasions and lacerations from impact with window cranks, radio knobs, door latches and dashboard components. Different types of pattern lacerations will also result from contact with the different kinds of automobile glass used in front and side windows (**Figs 7–9**). Analysis of the occupant's injuries should be undertaken by a physician/nurse trained in forensic medicine.

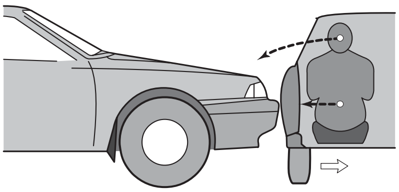


Figure 1 Occupant kinematics is defined as the movement of an occupant within a vehicle. An occupant will experience movement towards the point of impact. This movement is parallel and opposite from the principal direction of force.

Figure 2 A seatbelt contusion may provide valuable evidence as to the position of an occupant. This patient's contusion extends from the left lateral neck across the right breast and was the result of contact with the driver's side seatbelt.

Figure 3 (see color plate 2) Contact with deploying air bags will result in injuries to the occupants, including abrasions. This patient sustained superficial abrasions overlying the abdomen, secondary to air bag deployment. Such injuries can be matched to the vehicle's air bag.

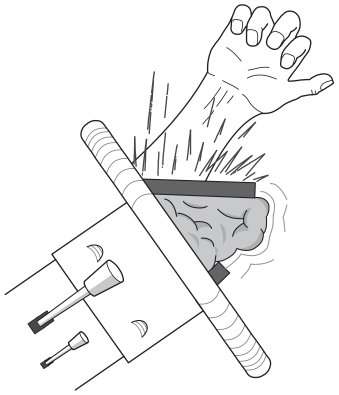


Figure 4 A diagrammatic representation of a forearm injury from contact with the deploying air bag module cover. This will give rise to the fracture pattern seen in Fig. 5.

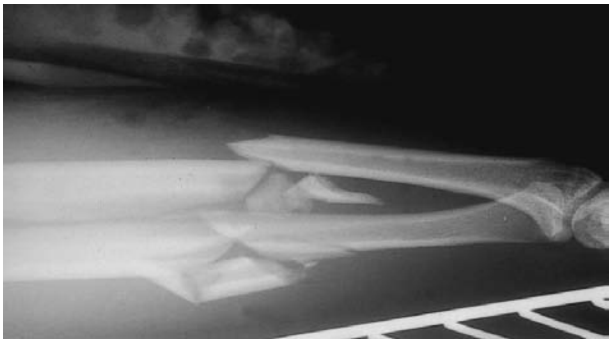


Figure 5 Contact with the deploying air bag module cover can give rise to pattern fractures of the upper extremities. This open comminuted fracture of the forearm indicates a bending type of fracture from contact with the air bag module cover.

Figure 6 (see color plate 3) A degloving injury with underlying fracture. This patient sustained an open fracture, with degloving of the forearm, secondary to contact with the deploying air bag. Matching this injury and the blood transferred to the air bag would assist in identifying the role of this patient.

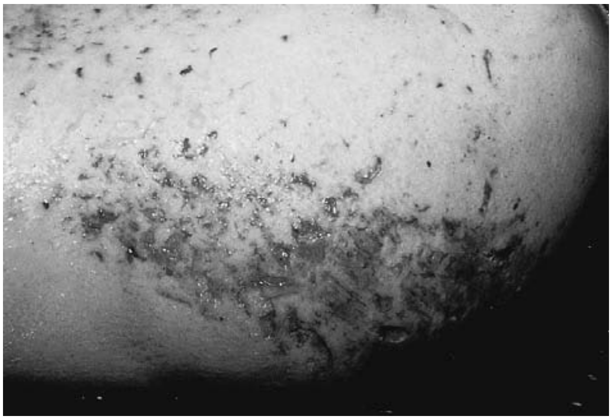


Figure 7 Contact with the tempered glass found in side and rear windows will impart a 'dicing' laceration to the skin.

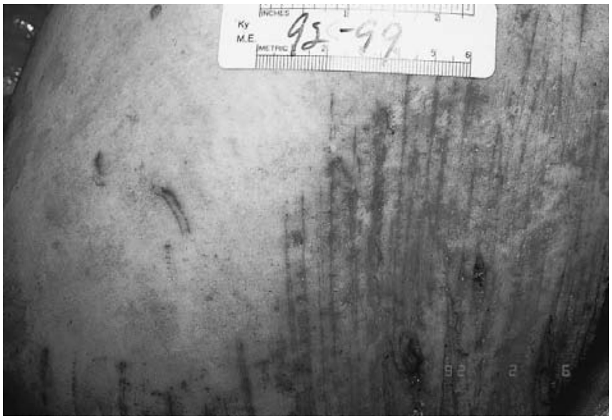


Figure 8 Sliding contact with tempered glass during occupant ejection will impart a characteristic superficial laceration to the skin. This is characterized by two parallel superficial lacerations 3–8 mm apart.

Figure 9 Contact with the laminated glass found in windshields will result in linear lacerations. These lacerations tend to be deeper than the dicing lacerations induced by tempered glass.

Trace Evidence

One of the most critical elements in the determination of the occupants' role in a motor vehicle collision is the collection of trace evidence from the victims and the vehicle. Special efforts must be made to collect clothing and biological standards from all vehicle occupants. The preservation of clothing, shoes and biological standards (including hair and blood) will be invaluable in determining an occupant's role. Examination of the soles of leather shoes may reveal the imprint of the accelerator or brake pedal, or the imprint of the leather floormat (**Figs 10** and **11**). The preservation of clothing will permit a forensic examiner to compare clothing fibers to those fibers transferred to vehicle components during the collision (**Fig. 12**). Fabric imprints may also be transferred to components within the vehicle (**Fig. 13**). Contact with the front windshield will frequently result in the cotransference of hair and tissue to the glass and of glass to the tissue (**Fig. 14**). Collection of this glass from a patient's wound can be matched with a particular window within the vehicle if a glass standard is collected from the vehicle involved.

Two types of glass are used in automobiles: laminated, used in windshields; and tempered, used in the side and rear windows. Each of these types of glass will produce a pattern laceration unique to it. The windshield is composed of two layers of glass, laminated together, with a thin layer of clear plastic sandwiched between. This laminated glass will break into shards upon impact; wounds resulting from impact with laminated glass will be linear and incised (Fig. 9). The tempered or 'safety glass' is a single layer of glass that breaks into small cubes when fractured. Wounds resulting from impact with tempered glass will appear 'diced', punctate and rectangular in configuration (Figs 7 and 8).

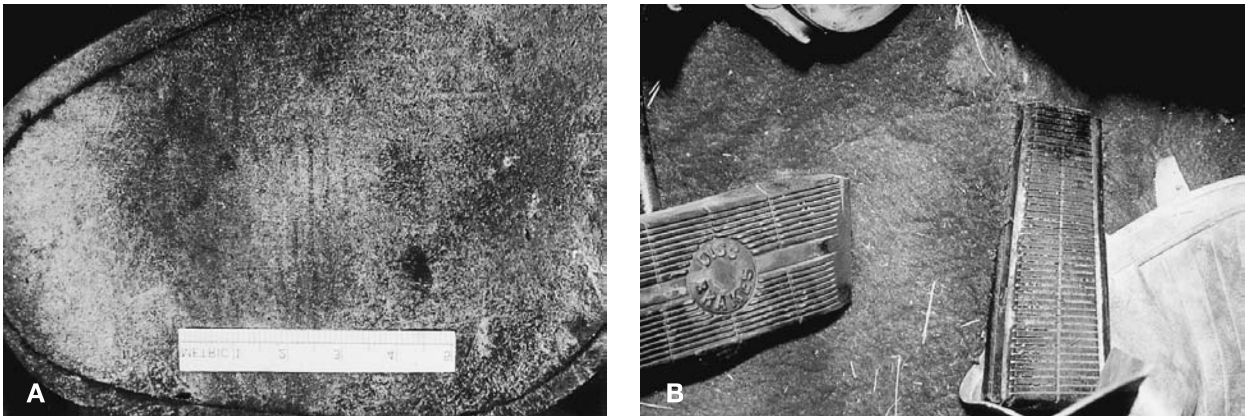


Figure 10 (A) Examination of leather-soled shoes may reveal the presence of a pedal imprint. **(B)** The shoe reflects an imprint from the vehicle's brake pedal and indicates that the wearer of the shoe was, in fact, the driver.

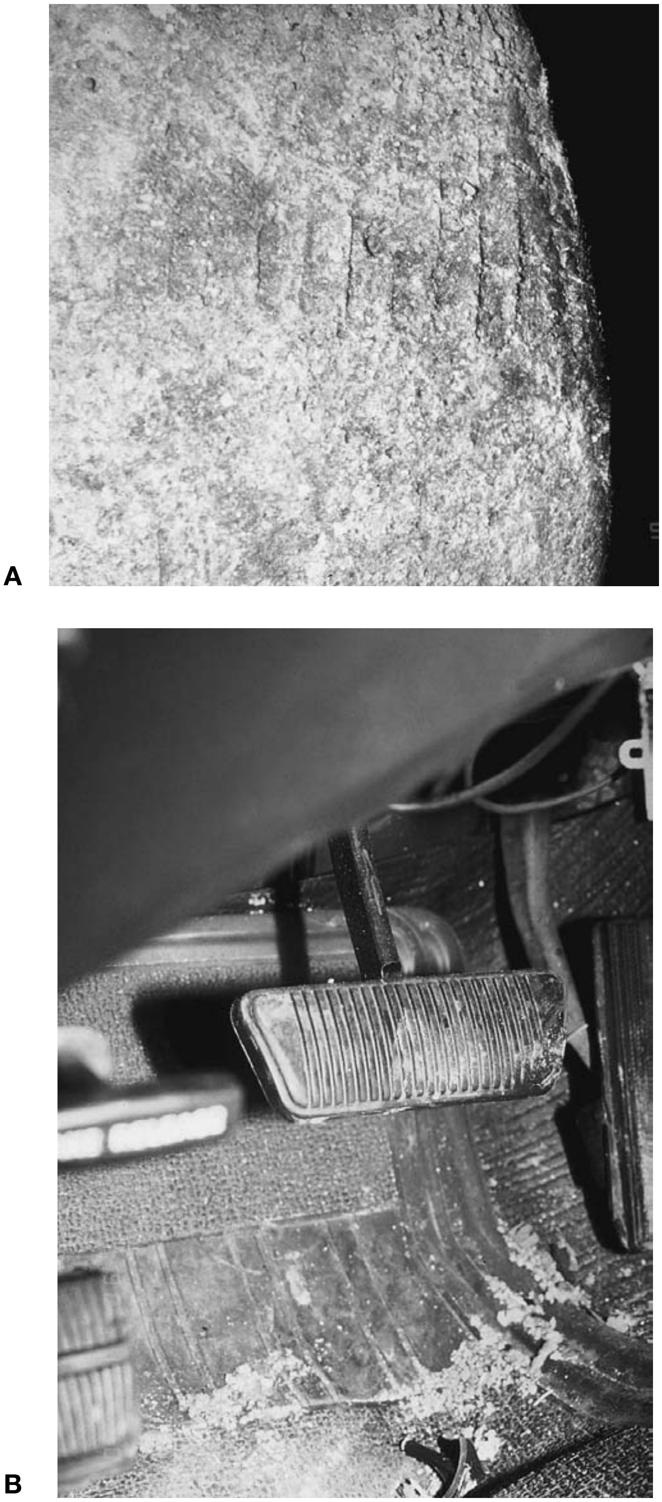


Figure 11 Examination of the dirt present on the sole of a boot **(A)** indicated the imprint of the brake pedal **(B)**.

The steering wheel may be damaged as the result of driver contact. The steering wheel in **Fig. 15A** displays rim damage from impact with the driver's mouth (Fig. 15B).

Close examination of both driver and passenger air bags will often reveal transference of trace evidence, including hair, tissue, make-up and blood to the bag's surface (**Fig. 16**). Occupant contact with the deploying air bags will also result in pattern injuries appearing on the patient (Figs 3–6). The air bag module cover

Figure 12 Fabric may be transferred from an occupant's clothing to components within the vehicle. Fabric from the passenger was transferred to the tempered glass in this rear window during her ejection.

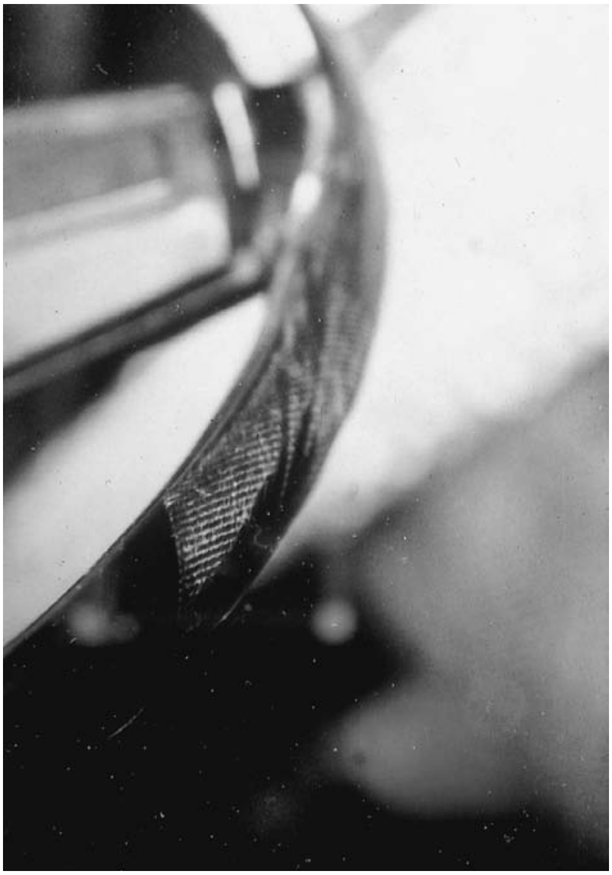


Figure 13 Forceful contact of clothing to vehicle components may impart a fabric transfer. A fabric imprint is noted on the steering wheel. Comparison of the weave pattern on the patient's pants to the pattern on the steering wheel confirmed the patient's role as the driver.

Figure 14 Occupant contact with the front windshield may reveal the transfer of hair, blood and tissue. This evidence can be matched to the standards obtained from the occupants.

may also yield valuable forensic evidence. Examination of the module cover may yield transferred tissue or an imprinted fabric pattern. Removal of all air bags from the vehicle should be done routinely in cases where occupant roles are in question, as they will yield a tremendous amount of forensic information.

Examination of the seat belt webbing and hardware may reveal the presence of loading marks if the occupant experienced a deceleration of the order of $25\,\mathrm{km\,h^{-1}}$ or greater (**Fig. 17**). These marks will confirm restraint usage and this evidence can be correlated with seat belt contusions on the occupants.

Sample Cases

Case 1

A 1956 Ford Thunderbird convertible was involved in a collision that resulted in the ejection of both front seat occupants (**Fig. 18**). One occupant sustained fatal injuries and the other sustained severe, closed head injuries. The surviving occupant had no recollection of the collision and was charged with vehicular homicide as he was the owner of the vehicle.

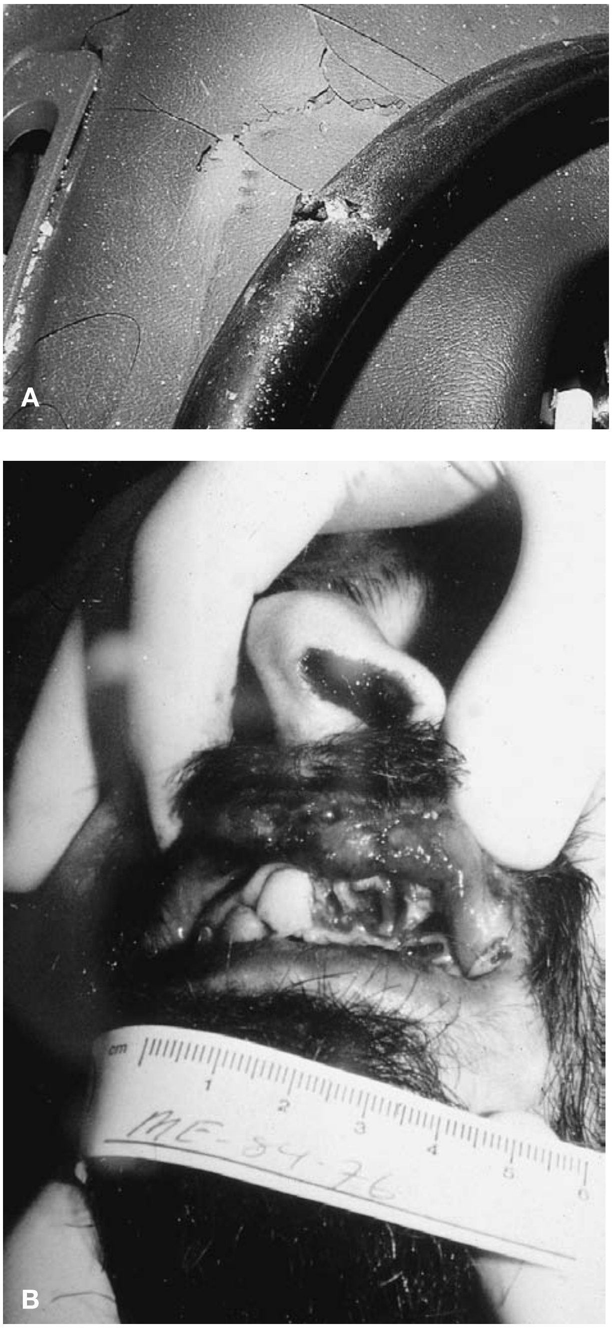


Figure 15 The presence of blood on the steering wheel in proximity to a divot of the steering wheel rim **(A)** from contact with the driver's teeth **(B)** confirmed the role of the occupant.

Figure 16 Examination of the air bag may also reveal the presence of transferred blood tissue, hair and makeup. Collection and preservation of a vehicle's air bag is obligatory if there is a question as to the role of the vehicle's occupants.

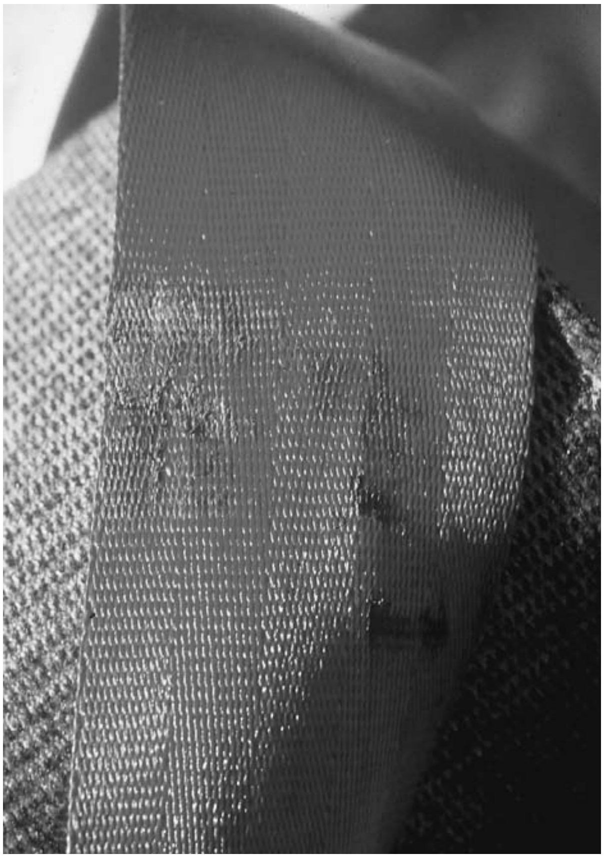


Figure 17 (see color plate 4) Examination of the seatbelt webbing and other hardware will be present if the belt sustained sufficient loading. Webbing and hardware should be inspected for evidence of transferred material or imprints from the webbing to the hardware.

The investigating agency failed to collect hair or blood standards from the deceased occupant, and only collected standards from the surviving occupant a year after the collision. The vehicle itself was not processed for trace evidence and the vehicle's interior was contaminated with vehicle components and scene evidence tape (**Fig. 19**). The vehicle was then left out, exposed to the elements, for over a year, without any concern for preservation of the vehicle's integrity. Despite these confounding variables and the fact that there was no witness to the collision, the vehicle owner was inappropriately convicted of vehicular homicide, as he was the registered owner of the vehicle.

Figure 18 A 1957 Ford Thunderbird convertible. Both front seat occupants were ejected, with one sustaining fatal injury. There were no witnesses to the collision and no evidence was collected from the vehicle by the investigating agencies. Despite the lack of investigative effort and the absence of physical evidence, the vehicle's owner was charged with vehicular homicide.

Figure 19 The interior of the vehicle (Fig. 18) was contaminated with vehicle components and scene evidence tape. The investigating agency took no steps to preserve or protect the integrity of trace evidence within the interior of the vehicle.

Case 2

A 1984 international semitrailer with two occupants rolled one-quarter turn to the right (towards the passenger door on to a guardrail), when exiting an interstate highway (**Fig. 20**). The surviving, bearded occupant, a 25-year-old man, was found partially ejected from the front windshield on the passenger side. The deceased, a 24-year-old bearded man, had been completely ejected and was found approximately 9 m behind the cab. The surviving occupant claimed that the deceased was the vehicle's operator. This contradicted the statement of a driver in a vehicle that had been passed moments before by that tractor trailer, who said the deceased was seen in the passenger seat.

Examination of the truck's interior revealed hair and tissue on the left side of a fan, situated on the ceiling in the middle of the cab (**Fig. 21**). Close

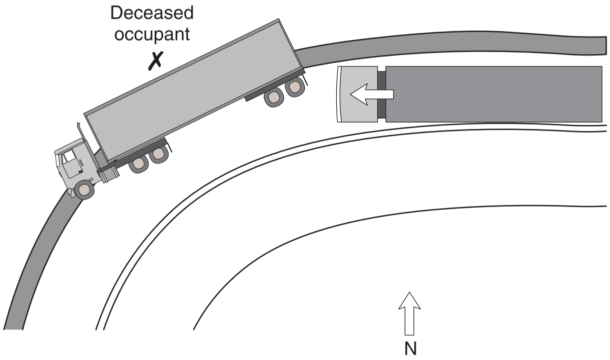


Figure 20 A semi-tractor trailer rotated 90° to the right, which resulted in the ejection of one front seat occupant. The surviving occupant stated that the deceased was the driver. Based on the statement of a passing witness, the surviving occupant was charged with being the driver.

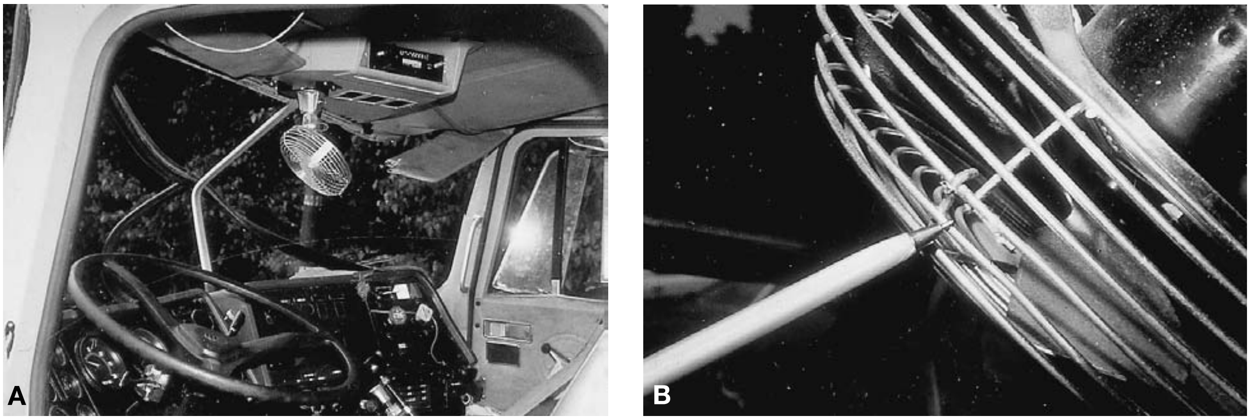


Figure 21 (A) There is a fan on the roof of the trailer cab (Fig. 20). **(B)** Close examination of the left side of the fan revealed the presence of trace evidence, including blood and tissue. Based upon a 90° rotation to the right, the occupant kinematics of the collision would dictate that the driver could impact the fan and it would be impossible for the passenger to make contact with the left side of the fan.

examination of the deceased's right forearm revealed a pattern abrasion on the extensor surface. The periodicity of the pattern abrasions matched the periodicity of the fan (**Fig. 22**). The conclusion rendered in this case supported the statement of the surviving occupant and contradicted the statement of the witness. This witness later stated that he may have been mistaken because both occupants of the truck were bearded.

Discussion

The ability to determine an occupant's role in a motor vehicle collision will depend upon the thoroughness of the examination performed at the incident scene, in the autopsy suite and at the hospital. It will also depend upon evidence recognition, collection and documentation. The following procedural recommendations are offered as a guide:

- Photographically document and determine, through interviews with witnesses and on-scene emergency medical personnel, the final resting places of vehicles and their occupants.
- Document and collect physical evidence of occupant contact and movement within the vehicle, including transferred hair, blood, tissue and fabric within the vehicle.
- Forensic autopsies must be performed on all deceased occupants of the vehicle.
- Clinical forensic examinations must be performed on all living occupants of the vehicle, including the photographic documentation and precise measurement of all pattern injuries.
- Atrauamatic areas of the body must also be documented.

Figure 22 At postmortem examination, a pattern abrasion was noted over the extensor surface of the right forearm. The periodicity of these abrasions matched the periodicity of the grill of the fan. Based on physical evidence and not eye-witness testimony, it was apparent that the deceased occupant was in fact the driver.

- Hair, blood and fabric standards must be collected from all occupants involved.
- Trace evidence found on the living and deceased, including glass, pieces of vehicle components and foreign bodies within wounds must be collected in the mortuary and in the emergency department. If surviving occupants are reluctant to provide hair, blood and fabric standards, a search warrant and court order must be obtained.
- Provide sufficient resources to the investigating agency to document and reconstruct the crash scenario and the vehicle dynamics adequately.
- Utilize a 'team' approach to these detailed and difficult investigations.

Table 1 lists resources used by some jurisdictions to successfully investigate and present cases of this nature.

Conclusions

The determination of an occupant's role in a motor vehicle crash may be a benign process if the occupant is trapped behind the wheel, or extremely complex if the occupants are ejected. It is critical that investigating agencies allocate adequate resources to these incidents, to insure that the correct role is assigned

Table 1 Investigation team

Players	*Role*
Accident reconstructionist	Determines vehicle dynamics, which is necessary to determine the occupant kinematics
Evidence technician	Processes vehicles for forensic evidence, collects evidence from deceased occupants at autopsy and collects evidence from living occupants in emergency department
Forensic physician/ nurse[a]	Evaluates occupants' kinematics and pattern injuries and trace evidence to determine role of each occupant. Performs autopsies on deceased occupants and clinical forensic examinations on living occupants
Prosecutor	Comes to scene and hospital to assist investigators in drawing up court orders and search warrants for collection of evidence. Ensures evidence is collected in proper fashion

[a] In some jurisdictions, clinical forensic examinations are performed by specialty-trained forensic physicians or nurse examiners.

to each occupant if criminal or civil charges are being considered. Developing and implementing a methodological procedure for the investigation and reconstruction of multioccupant fatal incidents is essential to the proper determination of the vehicle occupants' roles in the collision and their relative culpability.

See also: **Accident Investigation**: Motor Vehicle; Determination of Cause: Overview; Airbag Related Injuries and Deaths.

Further Reading

Auten J (1986) Why police officers don't investigate traffic collisions. *Law and Order* **34**.

Cooperstein IS (1984) Determination of cause of damage to an automobile. *Journal of Forensic Sciences* **29**:923–929.

Smock WS, Nichols GR, Fuller PM *et al*. (1989) The forensic pathologist and the determination of driver versus passenger in motor vehicle collisions. *American Journal of Forensic Medicine and Pathology* **10**:105–114.

Spitz WU (1993) The road traffic victim. In: Spitz WU and Fisher RS (eds) *Medicolegal Investigation of Death*, 3rd edn. Springfield, IL: Thomas.

Stone IC (1987) *Forensic Laboratory Support to Accident Reconstruction*, SAE 870427. Warrendale, PA: Society of Automotive Engineers.

Vankirk DJ (1984) A scientific approach to documenting evidence for accident reconstruction. *Journal of Forensic Sciences* **29**:8806–8815.

Motor Vehicle

S C Batterman, School of Engineering and Applied Science, and School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

S D Batterman, Consultants Associates Inc., Cherry Hill, NJ, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0391

Introduction

The purpose of this article is to provide an introduction to the fields of automobile accident reconstruction and biomechanics of injury. Automobile accident reconstruction is only a small subset of the general field of forensic engineering, which is defined as applying engineering principles towards the purposes of the law. Of necessity, the article will be limited in scope, as it cannot consider the contributions of all engineering disciplines to the field of forensic engineering. The reconstruction of an accident depends on physics and mathematics and falls naturally within the domain of the engineer. Engineering methodology can also be applied to the analysis of the mechanics of injury, some aspects of which are deeply rooted in medicine and biology. In this regard the last quarter of the twentieth century has seen enormous growth of the field of bioengineering with more than 60 undergraduate and graduate degree granting programs in the United States alone. Bioengineering, often referred to as biomedical engineering, combines and applies all the complexities of traditional engineering fields to the analysis of biological systems. Hence, only limited aspects of the application of engineering to injury investigation and analysis is considered here.

Attention in this article is focused on the reconstruction of automobile accidents, including the movement of occupants relative to a crashing vehicle (occupant kinematics), and some aspects of biomechanics and the mechanics of injury. The general underlying principles and methodology are discussed and a detailed illustrative example reconstructing an intersection collision is provided. A comprehensive bibliography provides more detailed information in accident reconstruction and biomechanics, as well as forensic engineering applications to other areas such as products liability. In products liability cases engineers are retained to analyze allegations of whether or not products are defective (either by design, manufacture or failure to warn/instruct) and if the defects are casually related to any injuries which may have been sustained. In automobile products liability cases, accident reconstructions and associated biomechanics analyses are often essential in determining if automobiles, and/or their components, are defective.

Accident reconstruction is not limited to automobile accidents and the methodology herein can be, and has been, applied to pedestrian accidents, industrial accidents, and aircraft accidents, to name a few. For example, the role of engineers in analyzing and reconstructing the terrorist bombing of PanAm 103 on 21 December 1988 over Lockerbie, Scotland is well-known and is an outstanding example of the role of forensic engineering in injury and death investigation.

Accident Reconstruction

General remarks

Accident reconstructions are *a posteriori* investigations of accidents or events with the purpose of trying

to determine how and what occurred during the event, consistent with the available data. In this sense, accident reconstructions are similar to human postmortem examinations or autopsies. Accident reconstruction is defined as the scientific process of analyzing an accident utilizing the available data and the appropriate natural or physical laws. The availability of data in combination with the knowledge of the reconstructionist are critical factors in determining the quality, accuracy and reliability of the reconstruction.

The general philosophy of accident reconstruction is one in which the reconstructionist generally works backwards in time and space. The starting point for the reconstruction is the final position of the vehicle(s) and/or accident aftermath proceeding backwards in time and space to the point(s) of impact and, if information is available, prior to the impact(s). The natural laws utilized in the reconstruction will be Newton's laws of motion. In addition, the reconstructionist must use the facts of the accident left at the scene and the physical data regarding the automobiles and the scene geometry. Typical information utilized in a reconstruction may include vehicle weights (masses) including those of passengers and cargo, point(s) of impact(s) on the roadway and on the vehicle(s), distances and trajectories moved post-impact, tire marks, gouge and scratch marks on the roadway, vehicle crush damage and damage locations, paint transfer, debris locations, body contact marks, location of blood and other body tissue, etc. The evidence left at the scene, or within a vehicle, or on a vehicle, or on a body are sometimes referred to as witness marks.

It is always best if the scene and vehicles are inspected as soon as possible after an accident. Not only do witness marks tend to disappear with time, but vehicles can be repaired or disposed of and the scene may change. Obviously, the quantity and quality of the available data will affect the accuracy and reliability of the reconstruction. Often the reconstructionist becomes involved in an investigation long after the accident occurred and may not have the opportunity to examine the scene and vehicles. Reliance will then have to be placed on investigation reports (police and others) and photographs. High quality photographs contain a large amount of useful information, some of which can be quantitatively extracted using principles of photogrammetry, the science of making measurements from photographs. In addition to dimensional information being obtained from photographs, the photographs can be digitally enhanced – not merely enlarged – to reveal significant details. Examples of information obtainable from photographs include scene dimensions, skid mark lengths, pot hole depths and vehicle crush patterns. Enhanced photographs have been used to show sharp edges of pot holes, tire defects, paint transfer marks, tire tracks, etc. Although physical inspections are always best, good quality photographs may also play a key role in understanding how an accident occurred and arriving at reliable and supportable scientific opinions.

Newton's laws of motion and related concepts

In 1687, Sir Isaac Newton stated the three natural laws of classical mechanics. These laws govern the motion of all bodies in the universe omitting relativistic effects, i.e. they are the operative laws for all problems when speeds do not approach the speed of light (more than 299 million m s^{-1} or about 186 000 miles s^{-1}). These laws, which apply to accident reconstruction, may be stated as follows.

First Law: every material body continues in its state of rest, or of constant velocity motion in a straight line, unless acted upon by external forces which cause it to change its state of rest or motion.

Second Law: the time rate of change of linear momentum of a particle (product of mass times velocity) is proportional to the external force acting on it and occurs in the direction of the force. An alternate form of this law, and the one that is most familiar, is that the resultant force ($\bar{\mathbf{F}}$) acting on the particle is equal to the product of its mass (m) times the acceleration ($\bar{\mathbf{a}}$) of the particle, or $\bar{\mathbf{F}} = m\bar{\mathbf{a}}$ (where vectors are indicated in boldface and a bar above the symbol).

Third Law To every action there is an equal and opposite reaction; or the mutual forces of two bodies acting upon each other are equal in magnitude and opposite in direction.

It is important to realize that there is a considerable amount of sophisticated philosophy embodied in Newton's laws which must be understood by the accident reconstructionist and which is critical to mention here. We note the following.

First, the laws are not scalar statements (which consider magnitude only) but are vector statements. This means that both the magnitude and direction of forces, velocities and accelerations must be considered. For example, speed and velocity are not interchangeable and constant speed is not the same as constant velocity. A vehicle traveling at constant speed in a circular path has a different velocity at each instant of time since the direction of the vehicle is constantly changing. In fact, the vehicle is indeed accelerating inward towards the center of curvature. However, a vehicle traveling in a straight line at a

constant speed is also traveling at a constant velocity and is not accelerating.

Second, 'motion' or 'rest' are relative terms when compared to a frame of reference. For accident reconstructions, the earth is a suitable 'fixed' frame of reference. However, for other applications such as rocket launchings, space flights and general astronomical considerations the earth would not be a suitable fixed reference frame and distant stars would then be used.

Third, Newton's laws mention forces without ever defining them. Force is an abstract concept and nobody has ever seen a force, i.e. we only feel the effects of forces. Forces may be defined as the actions of one body on another, and are classified as either contact forces (surface forces) or force at a distance, such as those due to gravity (weight) or magnetic effects. Again, we note that force is a vector quantity which has both magnitude and direction. It is further worth noting that in automobile accidents, apart from gravitational forces, forces on occupants only arise when the occupant contacts or is restrained by another surface such as a door or other vehicle structure, seat, seat belt, airbag, another person, etc. When this occurs the occupant is accelerated (decelerated) and Newton's second law must then be used to determine the forces acting on the occupant.

Fourth, Newton's laws are stated only for a particle, which is a mathematical idealization of a point with constant mass. It requires further rigorous treatment to extend Newton's laws to the motion of rigid or deformable bodies of finite dimensions as well as to the motion of the mass center of a body. Accident reconstructionists who lack an adequate educational background, or who are not engineers, may not appreciate this subtlety.

Fifth, the concept of units is embedded in Newton's second law. In the International System (SI) mass is given in kilograms (kg), acceleration in meters per second squared ($m\,s^{-2}$) and force is the derived unit and given in Newtons (N). One Newton is the force required to impart an acceleration of $1\,m\ s^{-2}$ to a mass of 1 kg and has the units of kg-m s^{-3}. In the US–British system of units, force is given in pounds (lb), acceleration in feet per second squared (ft s^{-2}) and mass is the derived unit given in slugs. One slug is equal to $1\,lb\text{-}s^2\,ft^{-1}$.

It is also worth noting that the second law introduces the concept of acceleration. Mathematically, acceleration of a particle $\bar{a}$ is the time rate of change of its velocity $\bar{v}$, or in the language of calculus, the first derivative of velocity with respect to time. For the purposes herein it is noted that acceleration can be considered the change in velocity divided by the short time interval of a crash over which the change in velocity occurs. This change in velocity, or Δv, is discussed further in the following sections.

Accident reconstruction calculations

The purposes of an accident reconstruction are to answer questions on the circumstances and contributing factors to how an accident occurred. A central feature of an engineering accident reconstruction is to obtain quantitative answers and this requires making calculations based on Newton's laws and the available data. Calculations can be performed by hand, or by computer, or both. Numerous computer programs for accident reconstruction are commercially available and it is not the primary purpose of this article to discuss these programs. Suffice it to say that computer programs are based on Newton's laws, as they must be, and many also contain special algorithms and techniques particular to automobile accident reconstruction. Some of the programs are so user friendly that often unqualified people, who do not understand the limitations of the programs or who do not have an adequate educational background, use the programs to obtain solutions that may not be valid for a particular accident. In addition, simulation programs are available which are not reconstruction programs. These programs, which contain many assumptions and limitations, proceed forward in time and space starting prior to impact, proceed to the vehicles impacting and then proceed to the post-impact rest positions of the vehicles. Answers obtained from these programs are not unique, i.e. several answers can be obtained for the same problem, and are very sensitive to the initial conditions and assumptions made by the reconstructionist.

Often, Newton's laws are cast in other forms for ease in performing calculations and solving a problem. Alternate forms of Newton's laws are the work-energy principle and the impulse-momentum principle. We emphasize that these principles are not separate laws of physics but are derived from Newton's second law and are mathematically integrated statements of the second law. The impulse-momentum principle is very useful in collision problems since the total external force acting on the system of colliding vehicles is zero at impact. This is equivalent to stating that the linear momentum of the vehicle system immediately before the impact is conserved and leads to the principle of conservation of linear momentum. We emphasize that conservation of linear momentum is also a vector principle where directions must be properly accounted for. It is also worth noting that although linear momentum is conserved at impact, the kinetic energy of the system of vehicles

is not, i.e. energy is dissipated in crushing the vehicles. However, as will be seen in the illustration, if sufficient information is available it is not necessary to use energy methods to solve a problem.

As an illustration of the concepts and calculation procedure discussed herein the common situation of right angle (90°) intersection collision of two vehicles is considered (**Fig. 1**). As shown, vehicle 1 proceeds through an intersection from left to right (in the positive x-direction) while vehicle 2 proceeds through the intersection (in the positive y-direction). Subscripts 1 and 2 are used to denote quantities for vehicles 1 and 2, respectively. A collision occurred and the vehicles came to rest in the positions shown. In this illustration each vehicle skidded prior to impact (shown by the lines behind the vehicles with lengths d_1 and d_2). Skidding means that the wheels locked up due to driver braking and the vehicles then slid, not rolled, to the point of impact. Investigation at the scene determined the point of impact (POI) to be as shown. This conclusion as to the location of the POI came from either an abrupt change in the direction of the skid marks, and/or gouges in the roadway surface and/or collision scrub marks from the tires. The location of vehicle debris on the roadway is in general not a good indicator of the POI since debris travels when it is dislodged from the vehicle and does not fall straight down to the roadway surface and stop. Impact speeds are designated with the superscript i, and post-impact speeds leaving the POI are designated by u_1 and u_2. The centers of gravity (or center of mass) of the vehicles moved the distances s_1 and s_2 to their rest positions as shown, while angles measured from the centers of gravity are also shown in the figure. The data shown in Fig. 1 are very complete and representative of an excellent on-scene investigation. In this case, the goal of the reconstruction will be to determine the speeds of the vehicles at impact ($v_1{}^i$ and $v_2{}^i$) and at the beginning of pre-impact skidding (v_1 and v_2). Although this example is relatively simple, it does illustrate many of the principles used in accident reconstruction.

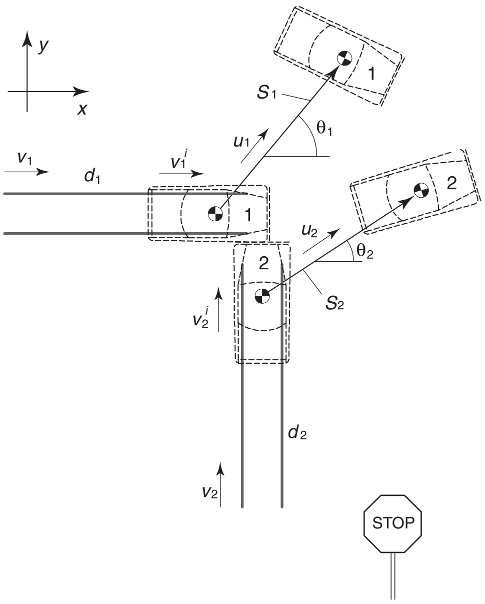


Figure 1 Intersection collision.

The basic equations used to solve this problem are all derived from Newton's second law. Since the philosophy is to work backwards from the points of rest to the POI, we first determine the post-impact speeds u_1, u_2 of the centers of gravity of the vehicles immediately after impact, i.e. at the point of vehicle separation. The reasonable assumption is made that the vehicles did not simply roll to their rest positions, but that the vehicles were displaced laterally and may have also rotated. This means that the tires were also moving laterally and essentially sliding over the road surface. For simplicity, the rotation of the vehicles to their rest positions is not considered in this example.

Noting that the centers of gravity of the vehicles moved distances s_1 and s_2 to rest, using Newton's second law it can be shown that the post-impact speeds are then given by

$$u_1 = \sqrt{2\mu \mathbf{g} s_1}$$

$$u_2 = \sqrt{2\mu \mathbf{g} s_2}$$

where μ is the coefficient of friction between the tires and the roadway surface and **g** is the acceleration due to gravity. At the earth's surface, **g** is approximately equal to 9.8 $\mathrm{m\,s^{-2}}$ (32.2 $\mathrm{ft\,s^{-2}}$). The coefficient of friction μ is a nondimensional parameter which is a measure of the relative slip resistance of two surfaces in contact. It can be measured at the scene on the day of the accident, which is preferred, or ranges of coefficients can be obtained from published references for tires on various roadway surfaces under a variety of conditions if it was not measured. In this case the coefficient of friction was measured and determined to be 0.65, which represents a typical road surface in average condition.

Conservation of linear momentum then states

$$m_1\bar{\mathbf{v}}_1{}^i + m_2\bar{\mathbf{v}}_2{}^i = m_1\bar{\mathbf{u}}_1 + m_2\bar{\mathbf{u}}_2$$

where m_1 and m_2 are the respective vehicle masses and bars over the velocities are used to denote vectors.

By taking the components of linear momentum in the x and y directions we then obtain

$$x: \quad m_1 v_1{}^i = m_1 u_1 \cos\theta_1 + m_2 u_2 \cos\theta_2$$

or

$$v_1{}^i = u_1 \cos\theta_1 + \frac{m_2}{m_1} u_2 \cos\theta_2$$

$$y: \quad m_2 v_2{}^i = m_1 u_1 \sin\theta_1 + m_2 u_2 \sin\theta_2$$

or

$$v_2{}^i = \frac{m_1 u_1}{m_2} \sin\theta_1 + u_2 \sin\theta_2$$

Note that in this example application of conservation of linear momentum in the x and y directions immediately gave rise to the impact speeds, without the necessity of solving two simultaneous equations, since the collision occurred at a right angle.

Since the vehicles skidded to impact, again using Newton's second law, it can be shown that the vehicle speeds at the commencement of skidding are given by

$$v_1 = \sqrt{2\mu g d_1 + (v_1{}^i)^2}$$

$$v_2 = \sqrt{2\mu g d_2 + (v_2{}^i)^2}$$

To obtain numerical values, the following data are used. Vehicle 1: m_1 = 1750 kg (weight = 3860 lb), s_1 = 6.0 m (19.7 ft), d_1 = 26.0 m (85.3 ft), θ_1 = 48°; Vehicle 2: m_2 = 2100 kg (weight = 4630 lb), s_2 = 4.6 m (15.1 ft), d_2 = 7.6 m (24.9 ft), θ_2 = 36°. Substituting into the above equations then yields

$$u_1 = 8.7 \text{ m s}^{-1} (31.3 \text{ km h}^{-1})(19.5 \text{ mph})$$

$$u_2 = 7.6 \text{ m s}^{-1} (27.4 \text{ km h}^{-1})(17.0 \text{ mph})$$

$$v_1^i = 13.2 \text{ m s}^{-1} (47.5 \text{ km h}^{-1})(29.5 \text{ mph})$$

$$v_2^i = 9.9 \text{ m s}^{-1} (35.6 \text{ km h}^{-1})(22.1 \text{ mph})$$

$$v_1 = 22.5 \text{ m s}^{-1} (81.0 \text{ km h}^{-1})(50.3 \text{ mph})$$

$$v_2 = 14.0 \text{ m s}^{-1} (50.4 \text{ km h}^{-1})(31.3 \text{ mph})$$

where the answers were first calculated in m s^{-1} and then converted to km h^{-1} and mph.

Hypothetically, the results of the above analysis can be used as follows. Suppose driver 2 claimed that he stopped at the stop sign, located 10.0 m (32.8 ft) from the POI in the negative-y direction, as shown in Fig. 1, and further claimed that the reason for the accident is that driver 1, who had the right of way, was speeding through the intersection. The speed limit was posted at 80 km h^{-1} (50 mph). Examination of the numbers obtained from this analysis indicates that vehicle 1 was sensibly traveling at the speed limit when the brakes were locked up. Furthermore, vehicle 2 could not have obtained a speed of 50.4 km h^{-1} at the start of skidding if that driver had started from rest at the stop sign. In fact, for a two-wheel drive vehicle, driver 2 likely could not have reached a speed of 35.6 km h^{-1} at impact starting from the stop sign even if no pre-impact skidding occurred. We note this for two reasons. First, driver 2 may claim that the skid marks were not from his vehicle. Second, often there are no pre-impact skid marks in an accident and the momentum analysis would then be able to reliably determine impact speeds only. We also note that for cars equipped with antilock braking systems (ABS) pre-impact skidding will typically not occur. Based on the results of the analysis an accident reconstructionist would then opine to a reasonable degree of engineering and scientific certainty that vehicle 2 did not stop at the stop sign prior to proceeding into the intersection.

Another quantifiable parameter used in accident reconstruction is that of determining the change in velocity, not merely the change in speed, of a vehicle due to an impact. The change in velocity, which is referred to as Δv, is defined as the change in velocity of the vehicle from its immediate pre-impact velocity to its immediate postimpact velocity and is one measure of the crash severity. Δv is a quantity which is often correlated with the injury-producing potential of an accident. As acceleration is the time rate of change of velocity, Δv is strongly related to, and is a measure of, the resultant vehicle acceleration in an impact. This is significant in determining occupant motion.

It is illustrative to use the numbers obtained from the aforementioned intersection collision example to demonstrate the calculation of Δv. Since Δv is also a vector, it is generally incorrect to merely subtract the speeds (magnitude of the velocity) before and after impact to obtain Δv. Hence, it is incorrect to calculate Δv of vehicle 1 as 31.3 − 47.5 = −16.2 km h^{-1} (−10.1 mph) decrease. The correct calculation requires using the components of the velocity vector in the x and y directions as follows.

$$\Delta v_x = u_1 \cos\theta_1 - v_1{}^i = 31.3 \cos 48° - 47.5$$
$$= -26.6 \text{ km h}^{-1} (-16.5 \text{ mph}) \text{ decrease}$$

(points in negative x-direction)

$$\Delta v_y = u_1 \sin\theta_1 - 0 = 31.3 \sin 48°$$
$$= 23.3 \text{ km h}^{-1} (14.5 \text{ mph}) \text{ increase}$$

(points in positive y-direction)

Hence, the magnitude of Δv for vehicle 1 is:

$$\Delta v_i = \sqrt{(\Delta v_x)^2 + (\Delta v_y)^2}$$
$$= \sqrt{(-26.6)^2 + (23.3)^2}$$
$$= 35.4 \text{ km h}^{-1} \text{ (22.0 mph)}$$

Similarly, for vehicle 2 it would be incorrect to calculate Δv as $27.4 - 35.6 = -8.2\,\text{km/h}^{-1}$ (-5.1 mph) decrease. The proper calculation for vehicle 2 follows using the velocity components.

$$\Delta v_x = u_2 \cos\theta_2 - 0 = 27.4 \cos 36°$$
$$= 22.2 \text{ km h}^{-1}\text{(13.8 mph)}$$
increase
(points in positive x-direction)

$$\Delta v_y = u_2 \sin\theta_2 - v_2^i = 27.4 \sin 36° - 35.6$$
$$= -19.5 \text{ km h}^{-1}(-12.1 \text{ mph})$$
decrease
(points in negative y-direction)

with the magnitude of Δv for vehicle 2

$$\Delta v_2 = \sqrt{(22.2)^2 + (-19.5)^2}$$
$$= 29.5 \text{ km h}^{-1}\text{(18.4 mph)}$$

Note that the incorrect calculations seriously underestimate the magnitude of Δv and say nothing about its direction. It is emphasized again that Δv is a vector with x and y components given above and is shown in **Fig. 2** for each vehicle. As shown, the Δv values are those for the centers of gravity of the vehicles which are not the same as those of the occupants when vehicle rotation is considered. Furthermore, by Newton's third law the impact forces on the vehicles have to be equal and opposite and it is these impact forces which determine the vehicle accelerations by Newton's second law. Hence, the vehicle Δvs have to be oppositely directed, as shown, although their magnitudes will generally be unequal since the vehicles have different masses.

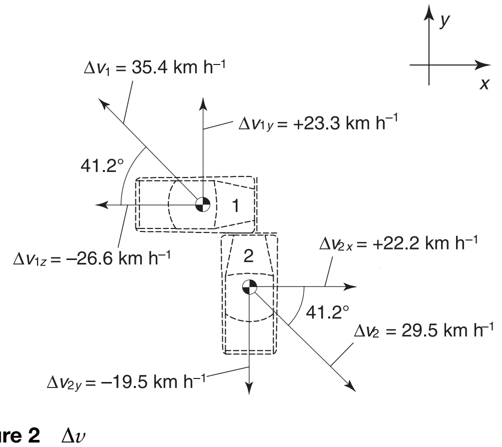


Figure 2 Δv

Before closing this calculation section it is important to mention another concept which often enters into accident reconstruction analyses. This is the concept of obtaining speeds from vehicle crush damage. It is noted immediately that only estimates to the change in velocity (Δv) can be obtained from an idealized crush analysis and not individual vehicle speeds, in general. Although there are special cases where vehicle speeds can be obtained, such as a collinear collision with one vehicle known to be stopped, an investigator cannot usually determine vehicle impact speeds from crush damage considerations alone. In fact, it cannot even be determined if the vehicle(s) were moving or stationary at impact. For example, a vehicle can be at rest (zero velocity) when it is impacted by another vehicle and obviously sustain crush damage.

Crush damage calculations are based on an idealized energy dissipation algorithm which is utilized in many computer programs. This algorithm contains assumptions on vehicle structural behavior, is based on an idealized crush pattern and also depends on certain measured properties of the vehicles known as stiffness coefficients, which are obtained from crash testing of vehicles. However, the damage algorithm is not a law of physics and its limitations and assumptions must be understood by accident reconstructionists prior to making damage-based calculations and drawing conclusions from them.

It is well-known that significant discrepancies can arise when the results of actual crash tests of vehicles are compared to the idealized energy dissipation crush analyses. If the data permits, a momentum calculation, which does not require any assumptions on vehicle damage or crush, is preferable to a damage only calculation. Note that in the example previously given, it was not necessary to use any aspect of vehicle crush or energy considerations in order to reconstruct the vehicle speeds. However, this should not be interpreted to mean that crush calculations are unimportant and should never be made. Sometimes the available data are insufficient to perform a momentum analysis to determine speeds. In this situation a crush analysis based on actual field measurements or on photogrammetric determinations of crush can provide useful information. In addition, a skilled and knowledgeable reconstructionist can often obtain bounds on the Δvs which may be significant in reconstructing an accident. Care and caution must always be used when interpreting the results of crush calculations.

Occupant kinematics and related concepts

The previous section dealt only with the reconstruction of vehicle speeds and no attention was paid to the occupants. Occupant kinematics is the term used to describe the motion of occupants relative to the vehicle in which they are riding, not the absolute motion of the occupant with respect to a fixed frame of reference. An occupant kinematics analysis may be a significant part of an accident reconstruction if an injury analysis is necessary. The complete determination of an occupant's motion in a crashing vehicle can be the most difficult part of an accident reconstruction and analysis.

The design purpose of vehicle seat belts is to restrain occupant motion in order to minimize or prevent occupant contact with interior vehicle structure. However, depending on both the magnitude and direction of Δv, i.e. the crash severity, occupants can still impact vehicle structures which will affect their occupant kinematics. In addition, it is the contact between the occupant and the vehicle structure, the so-called 'second collision', that can cause serious injury or death in an accident. The second collision causes an occupant to rapidly decelerate, and by Newton's second law, thus have high forces imposed on his/her body. Even properly worn and designed seat belts cannot prevent an occupant from impacting interior structure in all types of crashes. However, there is a dramatic difference in the protection afforded by different types of seat belt systems. For example, combination lap–shoulder belts are superior to lap only belts. Lap belts do not prevent upper torso jackknifing which can lead to serious injuries, even death, in relatively minor crashes. 'Lap belt syndrome' injuries, such as abdominal injuries and tension fractures (Chance fractures) of the lumbar spine can occur in frontal crashes with lap-belted occupants, typically seated in the rear of the vehicle. In a given accident, these same injuries would likely not have occurred if the vehicle manufacturer had provided lap–shoulder belts in the rear. Fortunately, rear seat lap only belts are rapidly vanishing from properly designed and manufactured vehicles, which is a trend that will continue into the twenty first century.

It is also worthwhile reviewing a few occupant kinematics scenarios for restrained and unrestrained occupants. In general, an occupant will tend to initially move towards the impact, i.e. towards the direction which is opposite of the vehicle Δv and the resultant vehicle acceleration. This is why care and caution was used in the illustrative example where Δv was properly calculated for both vehicles. Because of the directions of the vehicle Δvs (Fig. 2), occupants in vehicle I will tend to move forward and to their right whereas occupants in vehicle 2 will tend to move forward and to their left. In a purely frontal vehicle crash, occupants will tend to move forward relative to the vehicle. If unrestrained, front seat occupants can impact the steering wheel, dashboard and windshield whereas rear seat occupants can impact the backs of the front seats. After the initial motion into or towards vehicle structure the occupants will rebound or be deflected or impact each other and otherwise undergo a complex motion. Seat belts can limit the occupant excursions in the vehicle but, depending on the Δv of the vehicle, may not prevent interior vehicle contacts.

In a vehicle which is impacted in the rear, occupants will first tend to move backwards into their seats followed by a rebound phase towards forward structure. Unrestrained occupants can move forward into interior structure whereas restrained occupants will not rebound as far. However, prior to rebounding an occupant's head and neck may extend or hyperextend over the seat back and/or head restraint leading to vertebral column injuries typically located in the cervical or high thoracic spine. Occupant anthropometry is significant and tall people, greater than the 95th percentile in height, may be at high risk in vehicles where the head restraint cannot be adjusted high enough or may not even be adjustable. Although it is not discussed in detail here, low Δv rear end accidents also have the potential for causing cervical and lumbar spine injuries (such as bulging or herniated intervertebral discs), but these typically have a delayed onset. Another occurrence in rear impacts is the failure of seat backs, even in relatively minor crashes. When this occurs occupants can tumble into the rear of a vehicle (back seat or cargo area) where serious injuries can occur. It is not commonly appreciated that the seat is a part of proper occupant packaging and failure of the seat back, or the entire seat not remaining in place, can lead to serious injury or death for restrained and unrestrained occupants.

Lateral crashes can be particularly devastating since the nearside occupant immediately adjacent to the impact site on the vehicle may be at high risk of serious or fatal injury independent of the design of the vehicle. Initial occupant movement is towards the impact and it is not uncommon for the nearside occupant to have physical contact with the impacting vehicle or roadside object, such as a pole or a tree. Lap–shoulder belts would generally not be effective in preventing lateral movement in this type of crash. Padding may be effective in minimizing the effect of interior vehicle contacts and lateral air bags, which are now becoming available, will also offer some protection. A restrained far side occupant in a lateral crash will derive benefit from a lap–shoulder belt which will limit occupant excursions towards the

lateral impact site. However, an unrestrained far-side occupant can move across the vehicle into interior structures as well as into other occupants, which can cause injuries or death.

Detailed descriptions of occupant kinematics generally require the use of very complex computer programs. These programs are fundamentally based on Newton's laws but also require the human body to be modeled as an engineering structure. In addition, a large amount of vehicle information is required such as physical properties of restraint systems and those interior vehicle areas that can be contacted by occupants. These include, but are not limited to, padding, door structures, air bags, seat belts, dashboard, etc. It is also generally necessary to specify a large amount of input information in order to achieve an accurate solution. Changes in the input data can result in significant changes in the output and, hence, care must be exercised when interpreting the results of a complicated occupant kinematics computer-generated solution.

Biomechanics of Injuries

Biomechanics is a compound word which means the application of Newtonian mechanics to biological systems, including but not limited to the human body. The field of biomechanics is vast with applications beyond those considered herein in an automotive context. For example, to name a few areas, biomechanicians are involved in the design of prostheses, artificial organs, anthropomorphic test devices, bioinstrumentation as well as with the analysis of biological systems on the macroscale and the microscale, such as cellular mechanics. Biomechanics of injuries has already been mentioned several times herein. The remaining discussion provides a few additional concepts, but a complete discussion of injury biomechanics is beyond the scope of this article.

Injuries are often correlated with the vehicle Δv for both restrained and unrestrained occupants. Apart from what has already been mentioned, it should be noted that for unrestrained occupants, Δv physically corresponds to the velocity of the occupant relative to the vehicle. For example, if a vehicle is traveling at $50\,\mathrm{km\,h^{-1}}$ and crashes into a brick wall, unrestrained occupants will move forward at $50\,\mathrm{km\,h^{-1}}$ relative to the vehicle as the vehicle is being brought to a stop by the frontal crash.

Since Δv is a vector, injury correlations appear in the biomechanics literature as a function of Δv, seating position and the type of crash (frontal, lateral, rear, etc.). It must be emphasized that the injury correlations are statistical in nature, not absolute, and only attempt to quantify the probability of injury. Sometimes people walk away from a high Δv accident without permanent injuries, whereas others are seriously injured or killed in a low Δv accident. Furthermore, the severity of the injury is usually given in terms of the abbreviated injury scale (AIS) published by the Association for the Advancement of Automotive Medicine (AAAM). The AIS is based on anatomical injury at the time of the accident and does not score impairments or disabilities that may result from the injuries. **Table 1** gives the injury severity code as used in the AIS.

The concept of threshold injury criteria is also important to the understanding of injury biomechanics. Threshold injury criteria, along with human tolerance conditions, refer to those combinations of forces, moments, impact durations, stresses, strains, etc., which will result in traumatic injury. Caution is in order here since it is not reasonable to expect that a given injury criterion, if it exists at all, will apply across the entire spectrum of the human population. This is because biomaterials (human tissues) have a normal variability range in response to loading, and all people are not equal in their biomechanical response to impact. Human impact response can be, and is, a function of many factors including but not limited to age, gender, pre-existing conditions which may or may not be symptomatic, anthropometry (body size and measurements), etc.

Currently there are essentially three threshold injury criteria in common use. These are written into the Federal Motor Vehicle Safety Standards (FMVSS) which vehicle manufacturers must comply with by law. The FMVSS is an attempt to require a minimum level of safety and crashworthiness for all automobiles sold in the United States. The injury criteria refer to the head, femur and chest and are as follows.

The first, known as the head injury criterion (HIC), requires that a certain mathematical expression which is an integral of the resultant acceleration at the center of gravity of a test dummy's head in a frontal crash cannot exceed the numerical value of 1000 or else the standard has not been met and the vehicle fails to comply. There are many limitations to

Table 1 Injury severity scale

AIS	*Injury severity*
0	None
1	Minor
2	Moderate
3	Serious
4	Severe
5	Critical
6	Maximum (currently untreatable, fatal)

the HIC which are worth noting. The HIC does not even define or discriminate between types of head injuries which can range from subdural hematomas to diffuse axonal injuries to skull fractures, etc. In addition the use of a test dummy may be open to question when attempting to assess the effects of injury on live human beings; a live person may survive a HIC significantly greater than 1000 but may be killed in a crash when the HIC is significantly less than 1000.

The second criterion refers to the force measurement made in the femur of a test dummy during a frontal crash. The compressive force in the femur cannot exceed 10.0 kN (2250 lb) or else the standard has not been met and the vehicle fails to comply. This injury criterion, like all injury criteria, does not account for the wide and normal range of biomaterial variability. Compressive fracture loads can be considerably less than 10.0 kN for a significant portion of the population.

The third criterion refers to the resultant acceleration measured at the center of gravity of the thorax (chest) of a human surrogate test dummy. The criterion requires that this acceleration cannot exceed 60 times the acceleration due to gravity (60 **g**) for intervals whose cumulative duration is not more than 3 ms in a test crash. Depending on the test dummy used, another chest injury criterion which relates to chest deflection may also be required. As above, shortcomings in these criteria are that they do not distinguish between types of chest trauma (rib or sternum fractures, transected aorta, etc.) and do not account for normal population variability.

For completeness it is noted that other injury criteria have also been proposed and/or are undergoing intense research investigation. Biomechanics injury research is a vast and complex field which requires the collaborative efforts of engineers, physicians and other researchers in order to understand the response of the human body to impact. The brief discussion presented herein has only been intended as an introductory exposure to the field. Major advances in understanding the biomechanics of injury can be expected in the twenty-first century with forensic applications to automobile accident reconstruction.

See also: **Accident Investigation**: Rail; Airbag Related Injuries and Deaths; Determination of Cause: Overview; Tachographs; Driver Versus Passenger in Motor Vehicle Collisions. **Anthropology**: Skeletal Trauma.

Further Reading

Accident Reconstruction Technology Collection (1998) CD-ROM, Society of Auto Engineers.

Batterman SC (1990) Education of forensic engineers. *American Academy of Forensic Sciences (Newsletter)* **20**:(6).

Batterman SC and Batterman SD (1999) Forensic engineering. In: Caplan YH and Frank RS (eds) *Medicolegal Death Investigation, Treatises in the Forensic Sciences*, 2nd edn. Colorado Springs: The Forensic Sciences Foundation Press.

Beer FP and Johnston ER (1977) *Vector Mechanics for Engineers: Statics and Dynamics*. New York: McGraw-Hill.

Society of Auto Engineers (1993) *Biomechanics of Impact Injury and Injury Tolerances of the Head–neck Complex*, SAE PT-43. Warrendale PA: Society of Auto Engineers.

Society of Auto Engineers (1994) *Biomechanics of Impact Injury and Injury Tolerances of the Thorax–shoulder Complex*, SAE PT-45. Warrendale, PA: Society of Auto Engineers.

Bohan TL and Damask AC (eds) (1995) *Forensic Accident Investigation: Motor Vehicles*. Charlottesville: Michie Butterworth.

Brach RM (1991) *Mechanical Impact Dynamics: Rigid Body Collisions*. New York: Wiley.

Campbell KL (1974) Energy basis of collision severity. In: *Proceedings of the 3rd International Conference on Occupant Protection*, paper 740565. Troy, MI: Society of Auto Engineers.

Carper KL (1989) *Forensic Engineering*. Amsterdam: Elsevier.

US Dept of Transportation (undated) *CRASH 3 Technical manual*. Cambridge, MA: NHTSA.

Damask AC and Damask JN (1990) *Injury Causation Analyses: Case Studies and Data Sources*. Warrendale, PA: Michie.

Evans FG (1973) *Mechanical Properties of Bone*. Springfield, IL: Charles C Thomas.

Greenwood DT (1977) *Classical dynamics*. New York: McGraw-Hill.

Society of Auto Engineers (1986) *Human Tolerance to Impact Conditions as Related to Motor Vehicle Design*, SAE J885. Warrendale, PA: Society of Auto Engineers.

Hyzer WG (1989) Forensic photogrammetry. In: Carper KL (ed.) *Forensic Engineering*. Amsterdam: Elsevier.

Limpert R (1984) *Motor Vehicle Accident Reconstruction and Cause Analysis*, 2nd edn. Tokyo: Michie.

McElhaney JH, Roberts VL and Hilyard JF (1976) *Handbook of Human Tolerance*. Tokyo: Japan Auto Research Institute.

Mills PJ and Hobbs CA (1984) The probability of injury to car occupants in frontal and side impacts. In: *Proceedings of 28th Stapp Car Crash Conference*, pp. 223–235. Warrendale, PA: Society of Auto Engineers.

Nahum AN and Melvin J (1985) *The Biomechanics of Trauma*. New York: Appleton-Century-Crofts.

Newman JA (1980) Head injury criteria in automotive crash testing. In: *Proceedings of the 24th Stapp Car Crash Conference*, pp. 703–747. Warrendale, PA: Society of Auto Engineers.

Pike JA (1990) *Automotive Safety: Anatomy, Injury, Testing and Regulation*. Warrendale, PA: Society of Auto Engineers.

Smith RA and Noga JT (1982) *Accuracy and Sensitivity of CRASH*, paper 821169. Warrendale, PA: Society of Auto Engineers.
Stapp Car Crash Conference Proceedings (1998) CD-ROM. Society of Auto Engineers.
Association of Advanced Auto Medicine (1990) *The Abbreviated Injury Scale*, 1990 revision. Des Plaines, IL: Association of Advanced Auto Medicine.
Thorpe JF and Middendorf WH (1979) *What Every Engineer Should Know About Product Liability*. New York: Marcel Dekker.
Weinstein AS, Twerski AD, Piehler HR and Donaher WA (1978) *Products Liability and the Reasonably Safe Product: a Guide for Management, Design and Marketing*. New York: Wiley.
Yamada H (1973) *Strength of Biological Materials*. Huntington, NY: Krieger.
Yeh H and Abrams JI (1960) *Principles of Mechanics of Solids and Fluids; Particle and Rigid Body Mechanics*. New York: McGraw-Hill.

Rail

P Whent, Peter WHENT Associates, Braintree, Essex, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0394

Introduction

For a variety of reasons the investigation of a train crash is one of the most complex of investigations. It can involve tensions between senior staff of the different organizations as each would like to be in charge of the scene and to lead the enquiry. Wreckage can also be strewn across the entire railway possibly closing a main intercity artery. This in turn can place the investigators under some considerable pressure to allow the re-opening of the line at the earliest opportunity.

Any examiner called to the scene can be the world expert but may be employed by one of the interested organizations that may be at risk of being prosecuted at the end of the enquiry. These all add an extra dimension for the senior investigator in seeking an impartial examination to ascertain the truth about how the disaster was occasioned.

This type of enquiry is vastly different from that of a murder investigation launched by the police following a person being killed. For a murder inquiry an independent element is provided in the investigation with the police calling the national forensic scientists from one of the independent laboratories. The difference in a rail crash is that those who have expert knowledge of trains and railways are likely to be in current employment in the railways with one of the rail companies involved.

Other complications can arise where a search and rescue operation takes place with the Fire Service and other personnel walking over the scene disturbing or destroying evidence that may be later required by the investigators.

This article attempts to provide an overview of a rail crash investigation indicating which agencies may be involved in different parts of the world. It also stresses the importance of an impartial investigation as well as providing best practice in order to overcome potential difficulties.

Investigation of a Rail Crash

Train near-miss incidents and accidents occur daily throughout the world. When there are passenger trains involved and the crash results in a number of deaths or injuries it becomes international news. As soon as this happens the matter is raised to ministerial level in that country ensuring the fullest attention is paid to issues of safety and practice. The Minister may institute a Public Inquiry in addition to the other investigations into the disaster.

Criminal investigation

Whereas it is the duty of the Emergency Services to facilitate the rescue of people, it is normally the police who carry out an investigation to ascertain whether there is any individual or corporate body liable under the criminal code. The police usually notify the coroner and normally carry out inquiries for the coroner to enable legal identification of the deceased in order to allow the coroner to find and record the cause of death.

Safety investigation

In addition to any criminal inquiry safety factors must be examined to ensure steps are taken to prevent a recurrence. In the USA the National Transportation Safety Board (NTSB), an independent Federal agency, stands charged with investigating significant accidents on the railways.

In England and Wales the Health and Safety Executive (commonly called the Railway Inspectorate) will launch an investigation, as the Railway Inspectorate

will in Denmark. In India the Commission of Railway Safety plays a vital part, and to maintain its independence it reports to the Ministry of Civil Aviation, but has an overall responsibility for the investigation.

In by-gone days these mishaps would have been investigated by parties ranging from untrained persons with limited resources to large investigative teams of experts with almost unlimited resources. Today this has improved considerably and there is normally good cooperation between the investigating agencies.

Nevertheless it is most likely that every serious rail accident will be investigated by a number of different agencies all with slightly different aims but all centered on establishing the cause of the crash.

The NTSB has authority and complete discretion to designate other organizations as parties to the investigation, thereby allowing the involvement of those who have the expertise required in the investigation. Likewise in England and Wales the Railway Inspectorate can enlist the support of any other person. In either case should there be a criminal investigation it is carried out in the USA by the FBI and in England and Wales by the police, normally the British Transport Police.

The investigation principles

Two overriding principles should be uppermost in the mind of every investigator into the cause of a train crash. The first is remaining open minded and not jumping to early conclusions. Despite there being some evidence immediately available or an opinion of an expert how the crash occurred the cause must never be presumed until all of the related inquiries and investigations have been fully completed. The investigator must retain the view that anything could have caused the accident until that option has been eliminated.

The second principle is for the investigator to act impartially and to take an objective view of the investigation. This is a very important principle for it is clearly in the public interest to have an impartial investigation. Members of the public are entitled to expect the investigator to be an individual of high integrity who will act impartially, and if the occasion demands independently of the other parties who have a vested or pecuniary interest in the outcome. For the conclusions of the investigator to carry the most weight the investigation must be viewed as having been conducted both impartially and objectively.

Public interest

With the privatized railway companies existing today it is even more important in the public interest that an independent investigation is conducted. Otherwise a private railway company will be seen to investigate itself using its own staff. If this occurs it will attract cries of ‘whitewash’ for it contravenes the rule of law that no person should be a judge in their own cause. It is not to the benefit of the public who use the railways to have the rail companies investigate themselves.

In England preprivatization then the British Railways Board (BRB) alone would have an interest in any police or Railway Inspectorate investigation. This was evidenced when, following the Clapham train crash, the Board Chairman openly accepted responsibility on behalf of the BRB. Nowadays the various privatized businesses of the rail industry have an individual interest. In a train crash this can include those companies whose trains were in collision, the company who owns the infrastructure, the business responsible for signalling, a train lease company that leases the rail stock, the track maintenance company who maintain the track and others.

As other countries take the privatization road these difficulties will arise and need to be addressed. For on every occasion each separate company will want to know the cause of the crash and carry out some form of investigation if only to limit its potential pecuniary liability.

Aim of the investigation

The overall aim of any investigation must be to establish the truth and to conclude the real cause of the accident. For the law-enforcement agencies, however, their additional aim is to seek evidence to help determine any criminal liability of an individual or a company.

Over time the apportionment of criminal liability may well assist in the apportionment of civil liability and identify management errors or omissions, but safety alone is usually ancillary to the main thrust of the investigation by the police. It is more pertinent to the investigations by the Railway Inspectorate or Safety Commission that have specific responsibility for railway safety.

However, should the police identify matters that affect safety they should be formally reported to the appropriate Safety Agency.

Agencies involved

Most organizations have no responsibility for law enforcement and accordingly their aims will be different and numerous. They may wish to apportion blame, to protect themselves or their companies, to improve safety or to seek to identify funding difficulties by the government of the day. These competing demands can create tensions at the scene of the inci-

dent. In order to combat this there must, in the very early stages of the investigation process, be a hierarchical structure understood by all the agencies involved.

This is particularly relevant when the police and the Railway Inspectorate or Safety Commission are involved. There is a potential for one investigating agency to act without reference to the other. Clearly both agencies are likely to want access to the same exhibits, will require certain testing of apparatus and require an opportunity to enter the scene.

In England and Wales there is a protocol in existence agreed between the Crown Prosecuting Solicitor (CPS), the Association of Chief Police Officers in England and Wales (ACPO), and the Railway Inspectorate (HMRI). This allows for the Police Service to take the lead in a criminal investigation particularly when murder or manslaughter is suspected. Although the investigation can be led by the local police it is usually the Transport Police who investigate train accidents as they have the responsibility for policing the railways, as well as working closely on a daily basis with the Railway Inspectorate.

In other countries the decision as to who should lead the investigation has to be taken early. In the USA the NTSB will establish investigative groups unless a criminal act is established when the FBI will become the lead federal investigative agency with the NTSB then providing any requested support.

Agencies involved in the investigation

In any crash which causes a number of deaths there is every possibility that several agencies will launch investigations simultaneously. Staff from each organization will require access to the scene and will all be seeking the same evidence in order to establish the cause. They may include: the Police; HM Coroner; HM Railway Inspectorate (Safety Commission); Railtrack (owner of infrastructure); any interested rail business (i.e. the privatized train operating company); Government deciding to hold a Public Enquiry.

Police and Railway Inspectorate (or Safety Commission)

Police will seek evidence to show criminal culpability as well as establish the cause of death for the Coroner's Inquiry. The Railway Inspectorate or relevant safety commission will investigate to identify any infraction of the Health and Safety Legislation either by an individual or any company, as well as examine safety-related issues. Once identified, the Safety Agency has the authority to prosecute an individual or a company, issue mandatory directives to the company for immediate compliance or make recommendations for early or future improvement.

As a result there is considerable benefit in having a joint investigation involving the police and the Railway Inspectorate or Safety Commission. It is not desirable to carry out a joint investigation with the separate rail companies being involved as both the Police and Railway Inspectorate may be investigating the conduct of some of the employees of those companies. Such a multiagency investigation would be seen as neither impartial nor wholly independent and that would be against the public interest.

However, since privatization of the railways in England and Wales, occasions have arisen where there are several rail companies involved, making it physically impossible to allow their representatives to be at the scene together. Allowing only one or two companies access while excluding others would demonstrate unfairness and accordingly the companies are not normally allowed into the scene in the early stages of the investigation. They will be given access at an appropriate time with the consent of the senior investigator.

Some of these issues are highlighted in the two case studies that follow. The first one relates to one of the most serious train accidents in recent times in England: the Clapham Rail Crash of 12 December 1988. The second illustration is the German express train that crashed on 4 June 1998 at Eschede. This case is still under investigation.

Clapham crash

On 12 December 1988 at about 0810 h a collision occurred between a London-bound passenger train and a train of empty coaches. As a result 35 people died and nearly 500 people were injured.

This incident necessitated a full-scale emergency response to evacuate people trapped and injured. It precipitated investigations by the Police, the Railway Inspectorate, and British Rail and was followed by a Public Inquiry announced by the Minister of Transport.

The London Metropolitan Police and the British Transport Police (BTP) were at the scene. It was decided early on that the BTP would lead the investigation. They worked closely with the Railway Inspectorate providing them with a report into the crash. The Public Inquiry followed at a later stage after the completion of the investigations.

Initially the scene had to be preserved by the police to allow for experts in the railways to inspect the

site, the train, the carriages, the track and signaling systems in order to ascertain the cause.

The cause was an old wire that during re-wiring had not been disconnected, cut back nor secured from its old contact. This gave the driver of the passenger train a signal to proceed when it should have been a red signal.

This was a technological problem that required rail experts to examine the signaling system. It is unlikely that any one else would have found the problem. Once found it was invaluable evidence that had to be photographed *in situ*.

The investigations that ensued had to cover actions of staff actions of management in addressing whether there had been any act of omission or neglect on their part. In the final assessment although no one person was criminally convicted British Railways Board were heavily fined as proceedings were brought by the Railway Inspectorate for matters contravening safety.

The Public Inquiry covered all aspects of the incident and the final report carried 93 recommendations.

The train crash at Eschede in Germany

In this crash an express train alleged to be travelling in excess of 120 mph broke loose from the track which according to the news media resulted in an estimated 200 persons being injured and 100 people being killed.

In the early stages of the investigation similar decisions had to be made as in the Clapham crash and a Police Force was nominated to investigate. There was a full-scale rescue operation mounted and the train, signals and track had to be examined by rail experts.

As the crash happened near a road bridge much speculation arose about the possible cause being a car driven over the bridge or the bridge falling having suddenly collapsed. Both of these were found to be untrue, but it reinforces the importance of the investigator keeping an open mind until all investigations are complete. The investigation is still active and awaits a decision as to the cause, but the case illustrates many of the issues earlier mentioned.

Perhaps the most important aspect is all agencies working in cooperation.

Cooperation

At the 1998 International Transportation Safety Association meeting the Chairman of the Canada TSB told the meeting that international cooperation was absolutely essential. The sharing of best practice and lessons learned is vital.

Best Practice

Early identification of the lead investigator

This must be agreed at the earliest stage possible to ensure one focus.

Rail company tests

Where a rail company wish to carry out specific tests immediately before any possible deterioration in readings (temperatures or brake pressures) an early direct approach should be made to the senior investigator. The investigator can:

1. Allow access but impose conditions;
2. Agree to have the test carried out allowing a rail company representative to be present during the test;
3. Agree to have the test carried out by the investigating team and then provide the results to the rail company and other interested parties.

The investigator must be seen to be fair and open-handed to all parties. This includes rail companies, the infrastructure owner, the train driver and other staff or parties with a vested interest. The principle should be that if one party is allowed access then all should have access.

Preservation and collection of evidence at the scene

The scene of the accident is always likely to hold the best evidence to establish the cause. The scene must be preserved for the experts to examine. Despite any on-going rescue by the Fire Brigade the train cab must be fully preserved and protected by a physical presence.

There may be more than one scene, especially following a derailment. Each one must be treated as a separate scene, even to the extent of considering different examiners attending to ensure no crosscontamination.

Inspection of the driving cab

The positions of all levers and dials in the cab must be noted and recorded by the forensic investigator. It is essential to record both in note form and by video (or photographically) the position of all appliances, particularly those relating to braking, acceleration, direction of travel, and any audible warning systems.

In the more modern trains there are likely to be one or more data-recording devices. These should be recovered for specialist forensic examination. The data so recorded will cover train speed, the period of time the brakes were applied and much other vital information.

Independent element

It is in the public interest to insure there is an independent element present in the investigation. If the Police and the Railway Inspectorate are involved then the Inspectorate should agree the selection of every forensic examiner. They should either oversee the inspection making sure the examination is video filmed.

Minimum necessary equipment to be in possession of any forensic examiner

Each forensic examiner and evidence collector should have access to the following minimum equipment:

Video or camera
Hand-held tape recorder and cassettes
Tape measure
Containers for sample collecting
Paper or report forms
Pen

Note: Written entries should not be written on scraps of paper for this may be crucial evidence in any ensuing judicial hearing.

Causal Areas to be Examined

There are a number of generic areas of cause that will be addressed in this type of investigation. The question to pose is: Was the train crash occasioned by:

1. Human factors;
2. Mechanical problems or failures;
3. Technological problems (including computers) or failures.

The cause can fall into one of these areas or be a combination of all.

Human factors

The senior investigator will at an early stage consider whether there are any possible human factors involved in the cause of the crash. Failure to address this immediately may allow evidence to be lost forever. The evidence of any mechanical or technological failure will remain to be found at the scene but this is not so for any human factor. Failure to interview the train driver at the scene or at the hospital and request a sample of breath for alcohol analysis can mean that this evidence is never obtained.

Indeed one of the first questions to be asked, in order that the issue can be eliminated is 'Is there reasonable cause to suspect that any person was under the influence of alcohol or drugs?'

In England and Wales under the Transport and Works Act it is an offence for any operator to be on duty with more than the prescribed level of alcohol in their blood. This does not apply in every country throughout the world. In England and Wales many train drivers and operators often request the police officer at the scene to allow them to provide a sample for examination in order to prove that they were not under the influence of alcohol at the time of the incident. Samples may also need to be taken to enable an examination for drugs. Failure to consider this in the very early stages of the incident or at the scene may render it impossible at a later stage to show the condition of the operator at the time of the crash.

This process not only relates to the driver of a train but includes the train crew, staff in the signaling or operations center and any person carrying out repairs or maintenance on the railway line. In fact anyone who was acting as an operator.

Actions of staff Another aspect to examine is whether staff acted properly in accordance with company policy and standard operating procedures. Due regard must be paid to whether any have been careless, negligent or omitted to fully carry out their duties.

Actions of management The investigation should examine whether the senior management or board members of any company involved have fully met their responsibilities in respect of safety. The question to be addressed is whether they acted reasonably, fully discharging their duty of care to the traveling passengers and staff? If it can be shown that any death was occasioned by gross negligence then an offense of manslaughter or corporate manslaughter may result.

Trespass and vandalism The investigation must always address whether the crash was caused by an act of vandalism. Has there been an intentional act of vandalism such as altering the signal aspect or placing an obstruction on the railway line?

Obstruction of the railway line, trespass and vandalism is a very serious and prevalent problem for the railways, and must always be considered a potential cause of any rail crash. **Table 1** reveals the seriousness of the problem in England and Wales, with the

Table 1 Percentage of train accidents involving vandalism or trespass

Period	*No. of accidents*	*No. of accidents involving malicious action*	*%*
1991/2	960	163	17
1995/6	989	321	32

increase in the percentage of train accidents that involve trespass and vandalism.

A full and thorough inspection of the track is essential to allow the investigator to reject this as a possible cause. The investigators should aim to eliminate this cause within the first few hours of any crash.

The senior investigator will only be able to rule out human factors as a cause once satisfied that

1. every person, in the widest sense, involved in the operating and controlling of the train acted properly and in accordance with proper procedures;
2. no third party was involved in any trespass or vandalism related aspect;
3. managers and directors met their responsibilities on safety-related matters.

Mechanical and electrical systems

The various mechanical parts of all machinery must be examined, and this includes a detailed examination of the train and carriages. The examiner must be satisfied that all moving parts had been properly maintained and were in good working order and that all mechanical items were of an acceptable standard. The following type of inspections need to be carried out.

The tracing and testing of each electrical circuit
The tracing and testing of hydraulic or pneumatic braking systems
Examination of all electrical wiring and circuitry
Examination for engine wear or component wear
Examination of lubricating oils to identify any contaminants
Examination of wheels and other moving parts for metal fatigue
Examination of wheel bearings

The senior investigator often lacks this special knowledge. For that reason it is beneficial to have a constant liaison between the senior investigator and the forensic examiners. The senior investigator may wish to make parallel investigations into the maintenance routine to ensure that proper and regular maintenance systems were in place.

Where any metal fatigue, corrosion of metal or rotting of wood is found further tests must be carried out to establish whether this contributed to the cause and whether steps should be taken immediately to address these problems on similar trains currently in use.

Some of these items may be tested to destruction and although it is of benefit to the investigation to have an independent examination it is good practice to offer the other interested parties the opportunity of sending their own examiner to observe the test. On the other hand where the item will not be completely destroyed in the examination the other examiners can be allowed to examine it themselves.

An expert must examine the signaling system in operation to insure the signaling system used and where possible these tests should be video recorded.

Technological systems or failures

In the future, trains are all likely to be computer driven, with the driver only present to override the system for any emergency. The Docklands Light Railway operating in London is a current example where computer-driven trains operate without drivers.

Computers are used for the planning and routing of trains and in the efficient signaling from the railway signaling centers. Experts must be identified and sent to the appropriate control center to take possession of the computer tapes for examination.

These tapes will need to be examined and the data interpreted for the investigator. There must be a systems examination to ensure the computer was operating properly and that data relevant to this train crash were correctly stored in the system.

Witnesses and Injured

In any train crash there are likely to be witnesses to the incident. These may be members of the train crew, passengers, people who are injured, passers-by or local residents who saw or heard the sound of the crash. They must all be interviewed.

If the witness had been injured the full extent of the injuries must be recorded for this can help to estimate the speed of the train prior to impact. If any one is killed in such a disaster a post-mortem examination must be held and full records taken.

See also: **Accident Investigation**: Determination of Cause: Overview. **Crime-scene Investigation and Examination**: Recovery of Human Remains; Major Incident Scene Management. **Health and Safety**: Including Risk Assessment.

Further Reading

British Transport Police (1996) *Annual Report of Chief Constable British Transport Police*. London.

Ferry TS (1988) *Modern Accident Investigation and Analysis*. New York: Wiley.

Health and Safety Executive (1998) *Work Related Deaths*. London: HSE.

Health and Safety Executive (1998) *Annual Report of HSE (Railway Inspectorate)*. London: HSE.

Hidden A, QC (1989) *Report of the Investigation into the Clapham Junction Railway Accident*. London: HMSO.

Home Office (1994) *Dealing with Disaster*. London: HMSO.

Tachographs

R Lambourn, Forensic Science Service, London, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0482

Introduction

The tachograph is a device fitted to motor vehicles which makes a recording on a paper chart of the speeds traveled and distances driven, together with details of the driver's periods of work and rest. It is most widely used in the countries of the European Union, where it is a requirement in most larger goods and passenger-carrying vehicles.

This article lays particular emphasis on the instrument as it is specified in EU Regulations, although the principles of chart analysis and the use of tachograph data described here are applicable to all types of machine. In any case, the user of tachograph data must be aware of the characteristics of the particular make and model from which it comes in order to appreciate the amount of information available and its limitations. The present range of models is large and new instruments appear from time to time; this article cannot therefore cover every detail of every instrument which will be encountered.

The Forensic Use of Tachograph Data

Tachograph data has two forensic applications:

1. in road accident investigation, to determine speeds at and immediately before an incident, the rate of braking, and the manner of driving;
2. in general criminal investigation, to find the route traveled and the time that a vehicle was at a particular location.

The Tachograph Chart

Figure 1A and **1B** shows typical tachograph charts, and illustrate the information provided by the instrument. The shape and design are not specified by the EU regulations, but the chart pattern shown here has become the standard for the industry. Other designs do exist, but the instruments which require them are obsolescent.

The chart is a disc (123 mm in diameter) coated with a material which blackens under pressure. In the center is an area where the driver writes his name, the starting and finishing places for his day's driving, the date, vehicle number and odometer readings.

Printed round the edge of the chart, and repeated further in, is the time scale of 24 h. Between these is the field where the speed is recorded, usually with a maximum of 125 km h^{-1}. Further in is the work mode field, where the driver indicates whether he is driving, engaged on other work, or resting. The indication is made either by the track in which the recording line runs (Fig. 1a) or by the thickness of the line (Fig. 1b). It is sometimes possible to extract useful information from the mode of work line when the vehicle has been moving very slowly.

The innermost recording is of distance, shown by a zig-zag line. One stroke of this line is made during 5 km of travel; a complete V indicates 10 km of movement, whereas other distances will create partial strokes.

The Tachograph Instrument

All current models of tachograph combine their function with that of a speedometer, and are therefore designed to be mounted in a vehicle's dashboard. **Figures 2** and **3** show two modern instruments. Externally they have the usual speedometer features – a speed dial and an odometer display – together with a clock and knobs with which the driver and a colleague can indicate their modes of work. (Most instruments hold two charts to cater for there being two drivers crewing the vehicle.)

In the majority of instruments the chart lies behind the face, which is hinged to open downward to allow insertion and removal. **Figure 4** shows an open instrument with a chart in place. The spindle on which it is mounted rotates once in 24 hours. When the instrument is closed the chart bears against three styluses which move up and down to make the recordings. The spindle and the hole in the center of the chart are oval to ensure the correct orientation of the chart with respect to time.

With the instrument in Fig. 3 the charts are inserted through two slots in the base, from where they are automatically positioned against the recording styluses.

Tachographs may operate either mechanically or electronically. Mechanical types, in which a rotating cable from the vehicle gearbox drives a cup-and-magnet system, are now obsolescent. Typical models are the Kienzle 1311 and the Veeder Root 1111. Electronic types receive a train of pulses from a transducer at the gearbox, the frequency of which is translated as speed. There are many models, for example: Motometer EGK100, Jaeger G.50 and G.51, Kienzle 1314, 1318 and 1319, Veeder Root 1400, 8300 and 8400.

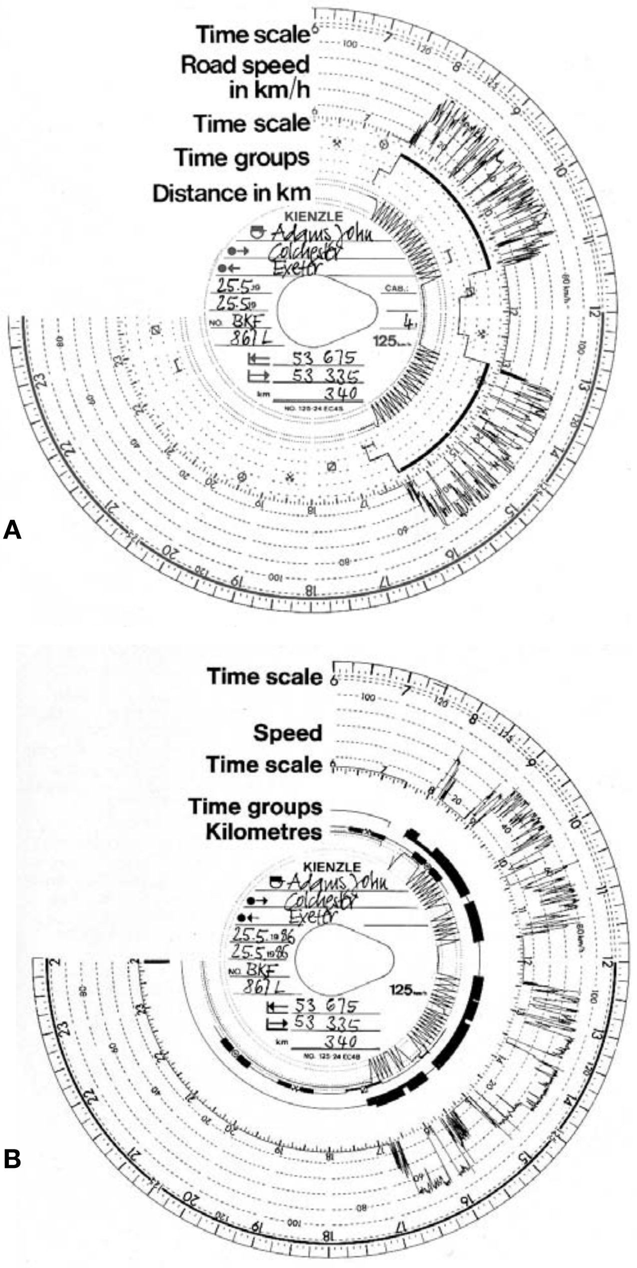


Figure 1 Tachograph chart showing recordings with (**A**) 'manual' time group recordings and (**B**) 'automatic' time group recordings. (*VDO Kienzle (UK) Ltd*)

The Principles of Chart Analysis

Figure 5 shows a schematic tachograph trace and the quantities to be measured in an analysis. The speed trace is thought of as a series of points connected by straight lines. The speed at each point (v_1, v_2, v_3) is measured, as are the time intervals between them (t_1, t_2, t_3). This provides a plot of speed against time, which can then be integrated to yield a plot of speed against distance. **Figure 6** shows a real tachograph trace, **Table 1** the tabulated analysis, and **Figure 7** the resulting plot. (This trace and analysis is used for the case example at the end of this article.)

There are various approaches to making these measurements. Taking the speed from the chart is relatively straightforward, and can be done from a photographic enlargement or with a traveling microscope. Allowance has to be made for the accuracy of the speed recordings. Measurement of the time intervals is, however, not at all easy. Because the chart rotates just once in 24 hours, one minute occupies only 0.25°, and in one second it turns through only 15

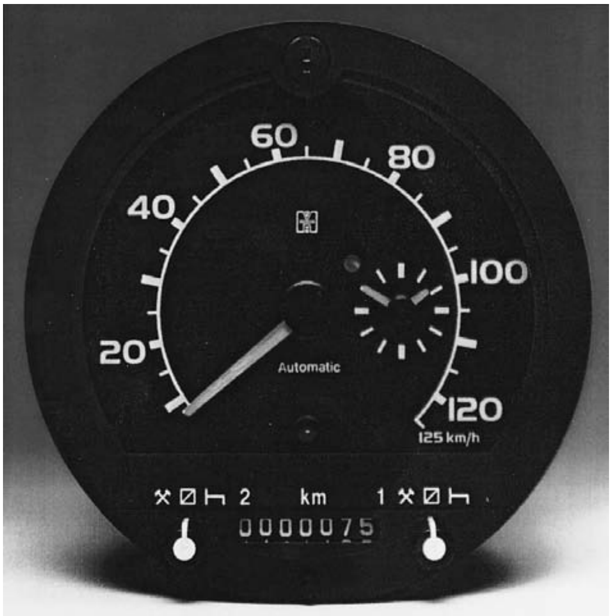


Figure 2 (see color plate 5) The face of a Veeder Root model 8400 tachograph. (*TVI Europe Ltd.*)

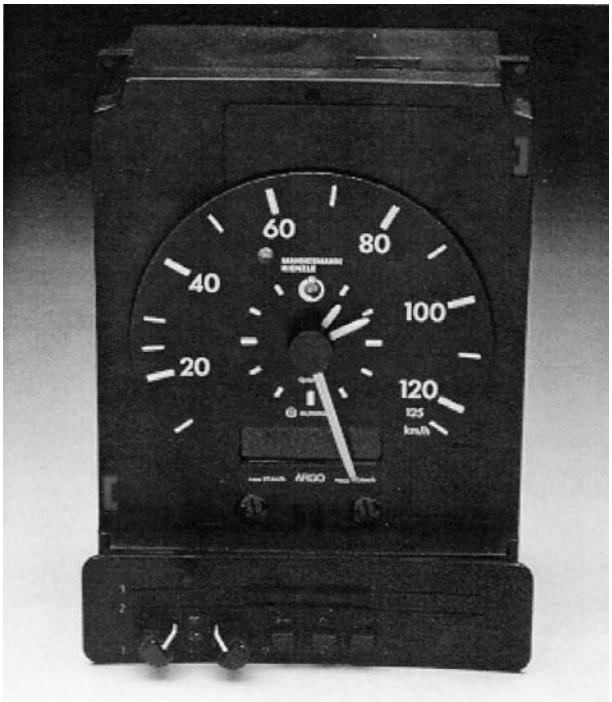


Figure 3 The face of Kienzle model 1319 technograph. The charts are inserted through the slots in the base. (*VDO Kienzle (UK) Ltd*)

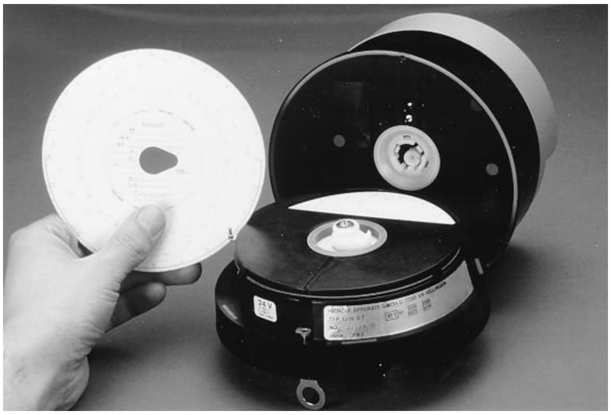


Figure 4 An opened Kienzle model 1318 tachograph showing the insertion of a chart. (*VDO Kienzle (UK) Ltd*)

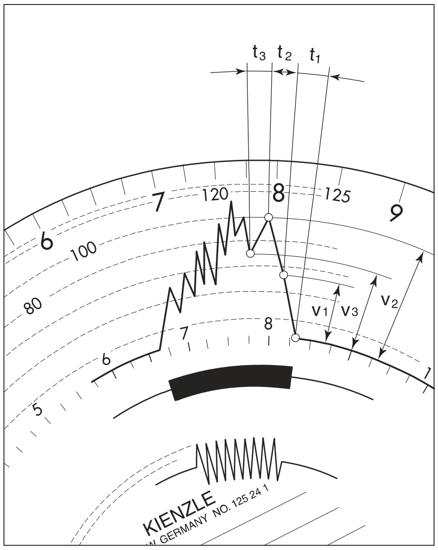


Figure 5 Schematic speed trace showing time intervals and speeds.

seconds of arc. It is usual to mount the chart beneath a microscope on a rotating table which is turned by a micrometer.

Figure 8 shows one such microscope, as produced by VDO Kienzle. A glass plate etched with a vertical cursor line, 3 μm wide, lies over the chart, and this combines with a horizontal line in one eyepiece to form a 'crosswire' in the field of view. Points for measurement are brought successively under the crosswire by rotating the chart table and moving it backwards and forwards, and transducers are attached to these two movements. Signals from the transducers, which are interpreted as time and speed, are fed to a microcomputer, where listings of these quantities and graphs of speed versus time and distance are produced.

It is very important to recognize when measuring time intervals that the line along which the recording stylus moved was unlikely to have passed exactly through either the center of rotation of the chart when mounted on the microscope or when mounted in the tachograph instrument, or through the geometric center of the chart itself. This is illustrated in exaggerated form in **Fig. 9**. If the vertical cursor is not arranged such that it lies along the line on which the stylus moved, any time intervals measured between

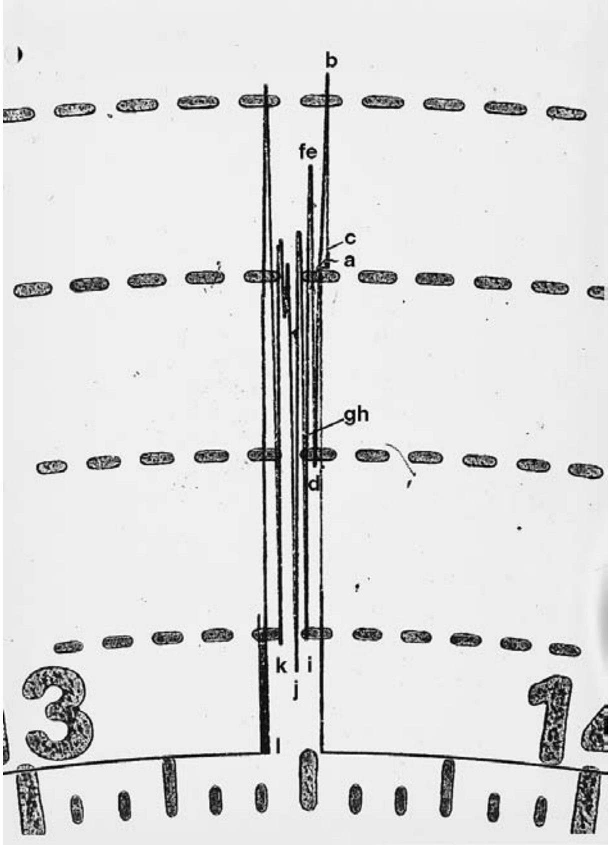


Figure 6 A real tachograph speed trace with analysis points.

Table 1 Data from tachograph chart shown in Fig. 6: the analysis points *a* ... *i* run backwards from the end of the driving

Point	*Speed* (*km h*$^{-1}$)	*Time* (*s*)	*Distance* (*m*)
a	–	0	0
b	81	0	0
c	61	32	630
d	38	67	1110
e	71	83	1355
f	71	87	1435
g	40	111	1805
h	39	116	1860
i	19	122	1905

points at different speeds will be in error. In the VDO Kienzle microscope the glass plate carrying the cursor line can be swivelled about a point at the outer edge of the chart, and this adjustment is used to set the line in what the analyst judges to be the correct position.

The Accuracy of the Speed Record

The EU regulations require that having been installed and calibrated, the tachograph record may not have an error in its speed figures greater than $\pm 6\,\text{km}\,\text{h}^{-1}$.

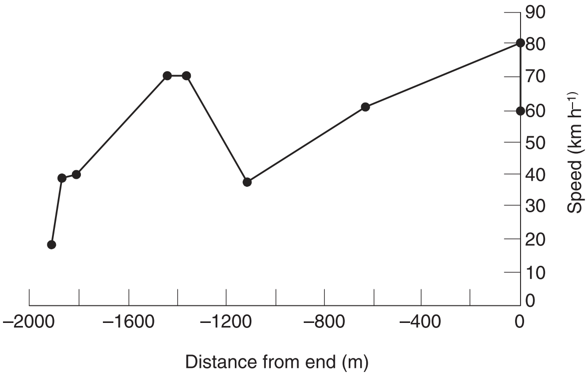


Figure 7 Plot of speed data from Fig. 6 against distance.

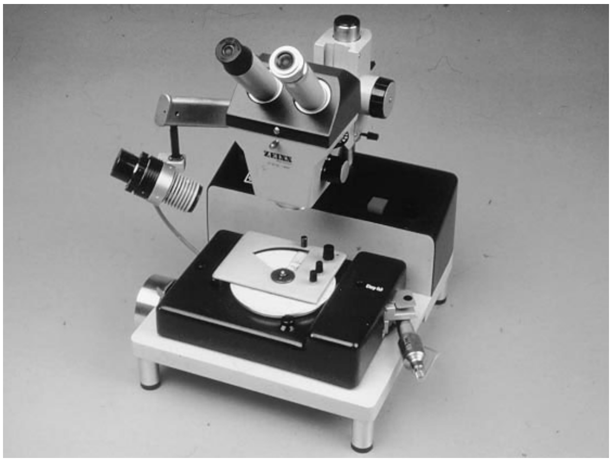


Figure 8 (see color plate 6) The Kienzle microscope with computer-linking modifications *(VDO Kienzle (UK) Ltd)*.

The analyst may be content to use this as a probable error figure, but the author's experience is that although the error is generally not greater than $\pm 3\,\text{km}\,\text{h}^{-1}$, there are occasional cases where it is substantially more than $\pm 6\,\text{km}\,\text{h}^{-1}$. It is therefore desirable that the calibration in each case be checked, and various methods have been devised for this.

Calibration checks on the tachograph installation

The most straightforward method of checking the accuracy of a tachograph system is to employ a 'rolling road' of the sort used when tachographs are installed in vehicles (**Fig. 10**). A chart is placed in the instrument, and the vehicle is then effectively driven through a range of known speeds. The chart will then show how each speed is recorded, from which the amount of error is found.

Some agencies find the rolling road method quite inconvenient, and instead use a method which requires no special facilities and which can be employed in any open space. The procedure is in two stages: first the signal sent from the vehicle's gearbox is measured

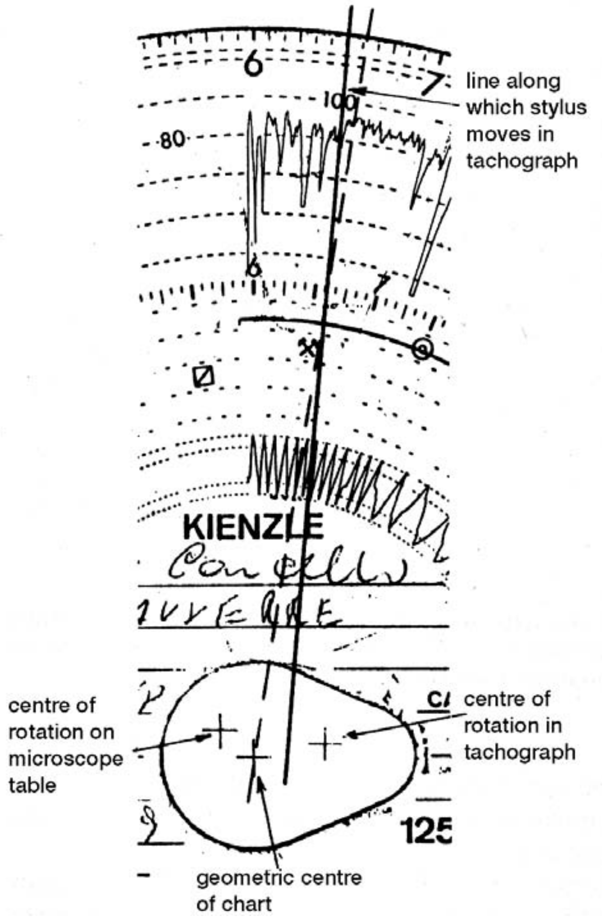


Figure 9 The need for an adjustable cursor when analyzing a chart.

Figure 10 Calibration of a tachograph installation on a Hartridge Rolling Road. (*VDO Kienzle (UK) Ltd*)

(either as pulses per kilometer in electronic systems or turns per kilometer in mechanical ones) in one of two ways, and then the way the tachograph instrument responds to a known set of input signals is found.

The first of the two ways of measuring the signal is called the 20-meter track, and is one which may actually be used by some installers. It is fully described in manuals produced by tachograph manufacturers for the guidance of their agents. It is appropriate for vehicles which can be moved slowly in a straight line; thus they need not be fully mobile, and may have, for example, accident damage to the steering gear or the engine coolant system. The test can even be performed by towing the vehicle.

In this method, the drive to the tachograph is either disconnected and a turn or pulse counter is attached to it, or, with later electronic models, a pulse counter is plugged into a socket in the head itself. The vehicle is then driven slowly on a straight and level surface for 20 m, the distance being measured exactly. From the number of pulses and the distance, the rate per meter is calculated: this is usually referred to as the *w* figure.

The second way is called the static method, and is appropriate for vehicles which cannot be moved, either through accident damage or through restriction of space. In this method a turn or pulse counter is attached in the same way as in the 20-metre track method. One of the driven wheels of the vehicle is then lifted off the ground and its circumference is measured with a tape; later the other driven wheels are also measured. The raised wheel is then rotated manually ten times and the resulting turn or pulse reading noted. From these data the *w* figure can be calculated, remembering to take account of the effect of the axle differential, since in general ten turns of a single wheel is equivalent to five turns of the whole axle.

If the accident damage is so great that neither of these methods can be used, the investigator may have to resort to examining gear trains and counting gear teeth.

Having determined *w*, the response of the tachograph instrument itself is found by driving it in steps through its speed range with a sequence of signals generated by a specially constructed test unit. These units are produced by the tachograph manufacturers for the use of their installation agents, and display either the mechanical rotation rate or the electronic pulse rate (in units of min^{-1}). A new tachograph chart is put in the instrument and a test chart similar to that in **Fig. 11** is produced.

The sequence of tests in Fig. 11 are as follows.

1. The instrument is taken up to its maximum speed ($140\,\text{km}\,\text{h}^{-1}$ in this example) and then very quickly back to zero to check the straightness of the recording stylus movement;
2. This is repeated three times;
3. Next the instrument is taken to an indicated speed of $60\,\text{km}\,\text{h}^{-1}$: at this speed the input figure (in min^{-1}) is known as the characteristic coefficient or *k* figure, and should be, in a correctly calibrated

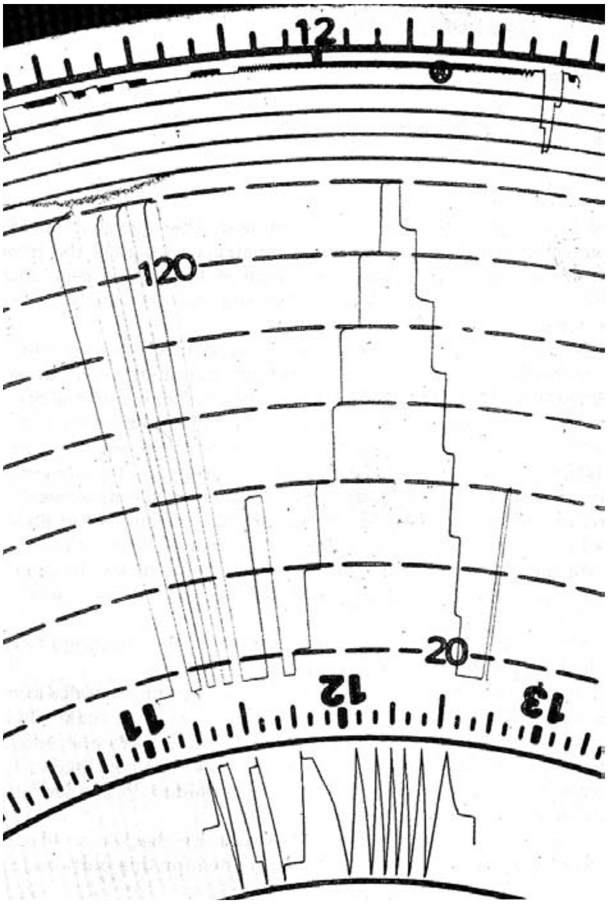


Figure 11 A typical head test chart.

system, approximately equal to the *w* figure of the vehicle;

4. The speed is taken from zero up to the maximum in 20 km h^{-1} steps, holding each step for two minutes, and then back down to zero at the same speeds between the steps (130 km h^{-1}, 110 km h^{-1} etc.);
5. Finally, the instrument is run at 60 km h^{-1} for as long as it takes 1 km to be recorded on its odometer, and the actual number of turns or pulses delivered is noted: this should equal the *k* figure; during the last test the activity mode switch is also changed through its four settings.

The information from part (4) of the test can now be used, together with the *w* figure, to construct a graph of how the recording on the chart at any point relates to the true speed. This can be done manually or by using the software associated with the VDO Kienzle computer-linked microscope.

The 'route-trace' method of calibration

An alternative to the direct methods of calibration is the so-called route-trace method. This can only be used when the route the vehicle has taken is known, and is particularly appropriate when the vehicle is unavailable for testing or has been so seriously damaged as to make the other methods impossible. The tachograph record can also be used to find the route that a vehicle has taken (see below). The calibration method is an inversion of this procedure, and briefly involves attempting to match the recorded pattern of driving to the known route. A system which is in calibration will give an immediate match, whereas one which is out of calibration will require an adjustment of its speeds by some factor before the match is found. Thus the calibration method is one of finding what this adjustment factor is.

Problems of Interpretation

Once the cursor on the analyzing microscope has been correctly aligned and the errors, if any, in the speed record have been determined, the chart can be analyzed. However, the limitations of the recording and the individual instruments must be recognized.

Time intervals

The clock in most tachograph instruments advances the chart in one-second steps and it is, therefore, not possible in principle to measure time intervals with a precision greater than ± 2 s. In any case the edge of the speed recording trace is insufficiently smooth for these steps to be resolved. With a good trace, a confidence of ± 2 s can probably be assumed, but in some instruments an unsteadiness will be evident which means that wider uncertainty must be accepted. It is found that although some manufacturers' instruments yield consistently good traces, other manufacturers produce tachographs which are invariably poor.

Figure 12 shows an example of a poor trace. The drive train from the clock to the chart is somewhat loose, and as well as allowing a generally unsteady speed trace as the stylus moves up and down, it permits the stylus, as it falls after rising to a maximum, to track back along the groove it has made on the chart instead of drawing a new line. This effectively drives the clock backwards for a few seconds (20 s being typical) until the slack is taken up, and only then will the stylus start to follow a new track.

Impacts to the vehicle

A severe impact to the vehicle will generally produce a disturbance of the tachograph recording traces (**Figs 6** and **13**). When such a disturbance is apparent on the speed trace, it is obviously an indication of the speed of the vehicle at the moment of impact and is an item of information easily read from the chart.

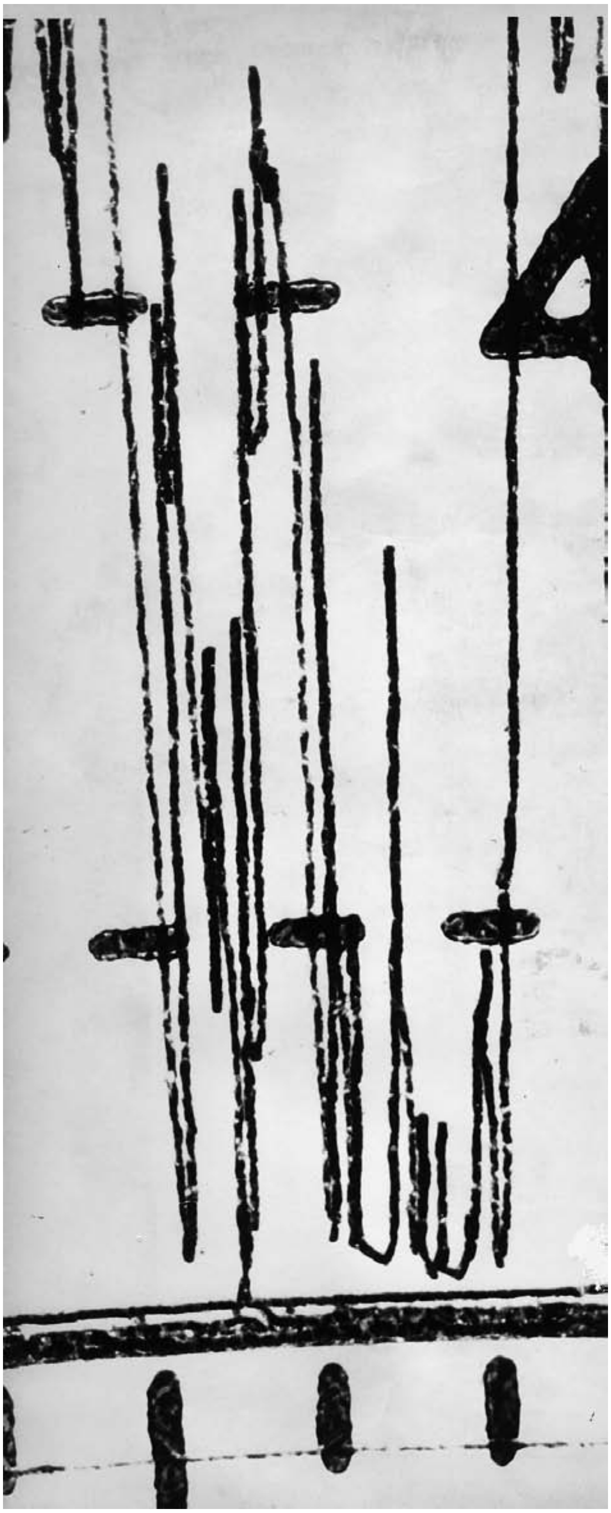


Figure 12 An unsteady trace from a Jaeger instrument.

However, some caution is needed before interpreting it. The disturbance is properly caused by a shock to the tachograph itself, and its magnitude therefore depends on the proximity of the impact to the instrument. A collision with a pedestrian, where the pedestrian meets the vehicle very close to the instrument panel, can cause a clear disturbance, whereas a severe impact from another vehicle to the rear of an articulated lorry may have little or no effect on the recording.

If heavy braking is taking place at the time of the impact, the speed at which the disturbance appears may be affected by other factors.

Finally, a severe impact will often derange the instrument for at least a few seconds, such that little or no reliance can be placed on any features which may immediately follow it.

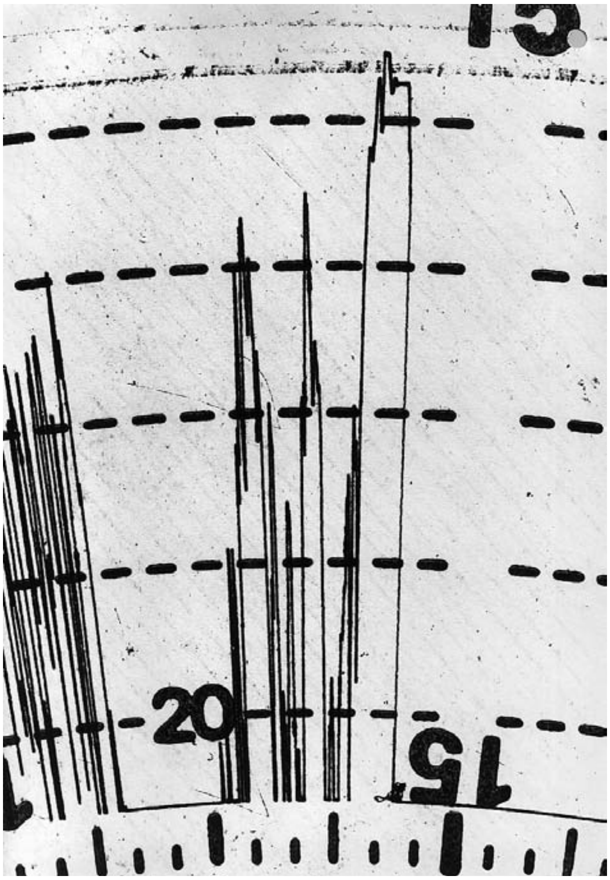


Figure 13 Chart showing lock-up and impact at zero speed.

Response of the instrument

The EU Regulations require tachograph instruments to be able to follow 'acceleration changes' of 2 m s^{-2}. Since during braking the deceleration of a vehicle may well be as much as 7 m s^{-2}, there is the possibility that the speed-recording stylus may not be able to keep up with the changing speed. This does not, in fact, appear to be a problem in Kienzle, Motometer, Jaeger and mechanical Veeder Root tachographs, but there is a difficulty with electronic Veeder Root instruments.

Earlier Veeder Root 1200 and 1400 series tacho-

graphs only just met the $2\,m\,s^{-2}$ requirement; early 8300 series models had a response of $2.5\,m\,s^{-2}$, whereas later 8300 series and all 8400 series models have a response of $5.7\,m\,s^{-2}$. The consequence of a response of 2 or $2.5\,m\,s^{-2}$ is that during heavy braking the speed being recorded at any moment will be greater than the true speed, and any event recorded during the braking, for example an impact, will be shown at too high a speed. It also means that it is not possible for an analyst to give an opinion as to whether a particular piece of braking was at a normal or an emergency rate.

A response of $5.7\,m\,s^{-2}$, however, is much closer to the braking rate which can be expected of a lorry in an emergency, and will generally allow a reasonable speed for an impact to be recorded.

An even higher response rate leads to a difficulty where there has been wheel slip during braking.

Tyre slip effects

The tachograph records the speed of rotation of the wheels to which it is connected: almost always the driven wheels of the vehicle. Therefore, if for some reason those wheels slip, i.e. rotate at a speed which does not correspond to the speed of the vehicle over the ground, the recording will be in error.

Wheel slip particularly occurs during heavy braking, notably when the driven wheels lock. An instrument with a fast response will react to wheel locking as though the vehicle had instantaneously come to a halt: the stylus will fall very quickly to its base position even though the vehicle is still traveling forward, and an impact may even appear to have been at zero speed. An example is shown in Fig. 13, where, because the vehicle has skidded, the prominent impact feature will certainly be at a level much lower than the true speed. The operation of an antilock braking system will also cause the speed of the wheels to be less than the road speed of the vehicle.

The opposite can occur if one of the driven wheels lifts off the road, allowing it to spin. This typically occurs as a vehicle starts to roll onto its side. In **Fig. 14** the speed at C would appear to be that at which the accident occurred. In fact the correct speed is at B, from which the trace has jumped to C as the wheels on one side of the lorry lifted from the road.

Determining at which point the true recording ends is not always a simple matter in these cases, but can often be resolved by using the route-trace method (see below).

Low speed behavior

Incidents at low speed cause particular problems in the interpretation of tachograph data because all instruments have a level below which they do not record speed. In all current models this is about $5\,km\,h^{-1}$, although in some older instruments it is somewhat higher, up to $16\,km\,h^{-1}$. It cannot therefore be said, for example, whether a vehicle which has slowed before entering a main road actually stopped or merely reduced its speed to, say, $5\,km\,h^{-1}$ before proceeding. Nor may it be possible to say whether or not during an apparent stationary period the vehicle was moved a small distance.

Some information can, however, be gained from the mode-of-work trace. Movement of the vehicle generates a thicker line, made by an oscillation, or an increased oscillation, of the recording stylus (see the examples in Fig. 1). A small gap in the thick line can be taken as a brief stationary period.

The threshold for this thick line depends on the model. Mechanical instruments will produce a noticeable movement in the stylus only when the vehicle has moved for about 20 m or more. Electronic instruments have a critical speed at which the line is made. The mode trace in Kienzle 1318 and 1319

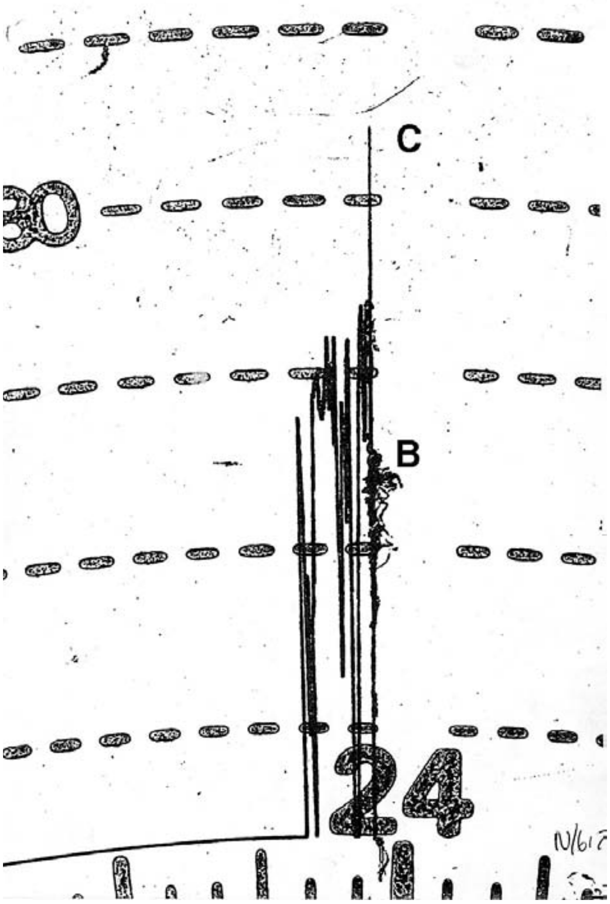


Figure 14 Spin-up in roll-over.

instruments, however, operates during driving in the same way as the zig-zag distance trace, making one stroke over a distance of 50 m.

Falsifications and Diagnostic Signals

Tachographs have always been prone to attempts by drivers to falsify the recordings, either to show lower speeds or shorter hours of work. Most falsifications are readily apparent on careful examination, for example by comparing the apparent average speed during a journey with the recorded distance and the time taken, or by comparing the total recorded distance for the day's driving with the difference in the hand-written odometer figures. **Figure 15** shows a commonly attempted falsification, where driving has stopped at 1736, the clock has been rewound, and driving has resumed, apparently earlier, at 1623.

The scope for such tampering is greatly increased with electronic instruments, and therefore various 'diagnostic' features are incorporated in them to show when certain irregularities occur. These depend on the individual models of instrument, and a full account of them is beyond the scope of this article. However, two examples are as follows.

If the pulse sender is disconnected in an attempt to make it appear that the vehicle is not being driven, the speed stylus will oscillate between zero and, typically, 30 km h^{-1} to make a broad band (**Fig. 16**).

If the electrical supply to the instrument is interrupted, in an attempt to halt recording for a period, then, on reconnection, the speed stylus will execute a characteristic movement, for example a full-scale deflection or a dip to zero speed (**Fig. 17**).

Other attempts at falsification which can be more difficult to detect include connecting the wires from the pulse sender to earth; reducing the voltage to the tachograph, such that the clock continues to operate but the other electronic systems fail; connecting certain spare terminals on the back of the instrument together, again to cause some of the electronic functions to fail.

Case Example

Figure 6 is an enlarged section of a tachograph recording made by a tourist bus which came into collision with a pedal cycle. The bus had been on a slip road, approaching a main dual-carriageway road. The pedal cyclist was travelling on the main road, and appears to have cut suddenly across the path of the

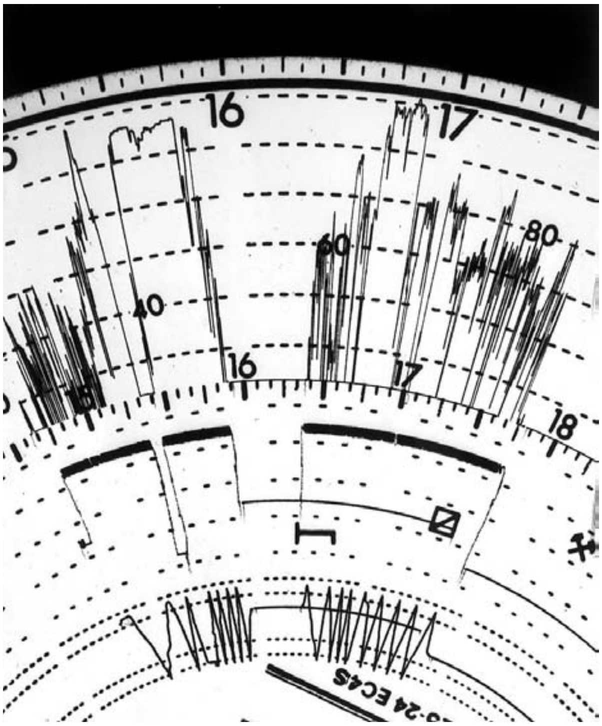


Figure 15 Fraudulent rewinding of tachograph clock. The overlapping traces show that the time was turned back from 1736 hrs to 1623 hrs.

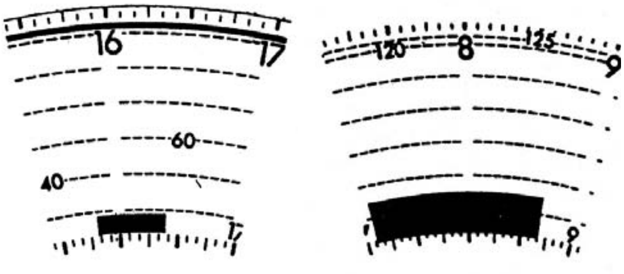


Figure 16 Broad trace generated when speed transducer is disconnected. *(VDO Kienzle (UK) Ltd)*

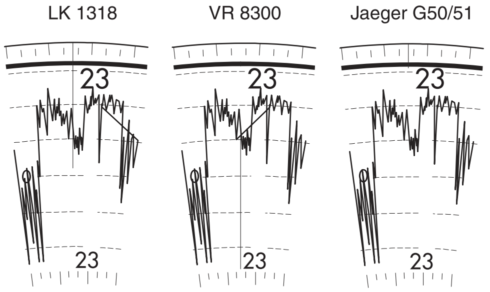


Figure 17 Diagnostic features produced by power interruption. (*VDO Kienzle (UK) Ltd)*

bus. The bus driver braked hard and swerved to his right. **Figure 18** shows the route the bus had taken to the scene of the accident.

The results of the detailed analysis, given in Table 1 and plotted in Fig. 7, indicate that (point *i*) about 120 s before the accident the speed of the bus was at a minimum of $19\,\mathrm{km\,h^{-1}}$, and that it was about 1900 m from the scene of the accident. Its speed then increased to a maximum (points *f* and *e*) of $71\,\mathrm{km\,h^{-1}}$ before falling to another minimum (point *d*) of $38\,\mathrm{km\,h^{-1}}$. It was then about 67 s and 1110 m from the end.

Again the speed increased, reaching its final maximum (point *b*) of $81\,\mathrm{km\,h^{-1}}$ immediately before the collision. There was then heavy braking to a sudden disturbance of the trace (point *a*), which on the chart is at an indicated speed of $60\,\mathrm{km\,h^{-1}}$. However, the time interval between *a* and *b* is too small to measure, and at the scene locked wheel skid marks were found from the rear wheels on one side of the vehicle. The rapid fall of the recording trace between the last two points has clearly occurred as a result of the locking of these driven wheels, which accounts for the immeasurably small time interval and also means that the figure of $60\,\mathrm{km\,h^{-1}}$ (because of the characteristics of the particular instrument) is probably significantly less than the speed at which the disturbance occurred.

The disturbance itself consists of a thickening of the trace, caused by a small shock to the stylus, followed by a step to the left (i.e. 'backwards' in time). The shock would have been caused by the impact at the front of the bus with the pedal cycle, whereas the sudden swerve would have brought about some sideways movement of the components of the tachograph instrument, to create the step to the left.

Figure 18 Section of Ordnance Survey map showing route of vehicle as determined from the tachograph recording in Figure 6. *(Crown copyright reserved)*

Route tracing

The two minimums of speed in Table 1 lie at distances from the end which, on an accurate map, can be readily identified with a roundabout (point *d*) and an intersection (point *i*). When the analysis is extended back to the start of the driving, two further minima and the start of the journey itself can be matched to a complex road junction and two more roundabouts (points *j*, *k* and *l* in Fig. 5). Some very slow driving before point *l* indicates that the bus started its driving from a place just before the roundabout.

In the context of a road accident, the matching of the trace to a known route serves to confirm the accuracy of the analysis process and of the calibration of the tachograph installation. If it had not been possible to check the calibration independently, this matching, with any necessary adjustment of the speed figures, could have been used as a calibration procedure in itself.

However, if it is imagined that, in different circumstances, the route of the vehicle was unknown and there was a need for it to be found, it can be seen that working back from the known end point (*a*) it would be possible to do so by comparing distances between speed minimums with distances on a map between road junctions. A typical case where this method is applied would be where a vehicle carrying a valuable load is stolen from a known location *X*, driven to an unknown place *Y* and unloaded, then driven to a third place *Z* and abandoned. From its tachograph record it may be possible to locate *Y* by tracing back from *Z* or forward from *X*. The success of this method depends on traffic conditions; in heavy traffic the vehicle may have had to stop so many times at places other then road junctions that following its route becomes impossible.

See also: **Accident Investigation**: Motor Vehicle; Determination of Cause: Overview.

Further Reading

EEC (1985) Council Regulation (EEC) no. 3821/85. *Official Journal of the European Communities* no. L370:8.

Lambourn RF (1985) The analysis of tachograph charts for road accident investigation. *Forensic Science International* **28**:181–199.

Lehmann H (1992) Die mikroskopische Diagrammscheibenauswertung für die Unfallrekonstruktion (Microscopic analysis of tachograph charts for accident reconstruction). *Verkehrsunfall und Fahrzeugtechnik* **1**:2–4.

Lowe D (1989) *The Tachograph*, 2nd edn. London: Kogan Page.

Needham P (1988) Tachograph chart computer-linked microscopic analysis. *Forensic Science International* **36**:211–218.

ACCREDITATION OF FORENSIC SCIENCE LABORATORIES

F P Smith, Department of Justice Sciences, The University of Alabama at Birmingham, Birmingham, AL, USA
D A Kidwell, Chemistry Division, US Naval Research Laboratory, Washington DC, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0399

Introduction

Historically, testimony in court was based on a combination of school learning and on-the-job training. Because both criteria can be quite variable, the confidence in the interpretation of results was also quite variable. Few generally agreed upon procedures

were available. The identification of a drug in a body fluid may rely upon spot tests, thin layer chromatography, high performance liquid chromatography, capillary electrophoresis (CE), fluorimetry, infrared, gas chromatography, gas chromatography/mass spectrometry, immunoassay, or a combination of the above. Consensus on the superiority of one method over another is difficult to obtain. For example, Professor Manfred R. Möller opined that it would be easier to get a group of forensic scientists to share the same toothbrush than to share the same method of analysis.

In other analytical laboratories, such as clinical laboratories upon whose results medical diagnoses are based, a similar pattern of public outcry resulted in the imposition of stiffer laboratory performance requirements. US Congressional testimony citing laboratory errors in medical diagnoses and lack of clear standards for detection of infectious diseases resulted in the passage of the Clinical Laboratories Improvement Act (CLIA) of 1988. This act established quality standards for all medical laboratory testing to ensure the accuracy, reliability and timeliness of patient test results regardless of where the test was performed. The final CLIA regulations were published on 28, February 1992 and have resulted in steadily improving quality, as measured by the precision and accuracy of results. The regulations specifically exempt research, National Laboratory Certification Progam (NIDA) drug testing (exempt because NLCP maintains its own accreditation program specific to the tasks at hand), and forensic science testing.

Laboratory accreditation implements guidelines to provide quality and standardized results. However, accreditation is not without detractions. Certainly, costs are incurred in any proficiency testing program, which must be weighed against the results and simplicity of the tests. In addition, standardization and proficiency testing mandated by the CLIA has resulted in the reduction and elimination of certain tests performed in physicians' offices; although the quality of results may have increased, the timeliness of results has certainly decreased. Recognition of the impact on family physicians has resulted in the US Congress reexamining the application of the CLIA to point-of-care testing.

Accreditation programs should examine both the specific procedure being followed and the general procedure for unknowns. Accreditation is more easily accomplished when several laboratories are performing the same basic tasks such as fingerprint analysis, routine drug and alcohol analysis, questioned document examination, hair and fiber comparisons, etc. Often in forensic cases, especially unique poisonings, the toxin or question to be answered may be so unusual that only general scientific principles can be evaluated for its detection rather than a set procedure. Certainly, for unusual toxins no accreditation program could reasonably include blind proficiency testing. Thus, the accreditation of forensic laboratories can be more lengthy and the procedure more difficult to define than that of general medical diagnosis.

Although older technology has its value, analytical technology is ever advancing to more sensitive, more reliable, and less error-prone techniques. Often, only after high visibility events, such as the Lindbergh kidnapping (1930s) (resulting in funding of FBI laboratory and local law enforcement laboratories), civil unrest (1960s) (with US Department of Justice funding of lab improvements, including proficiency testing), or more recently the OJ Simpson trial and FBI whistle blower scandal (with increased calls for better evidence collection procedures and more money for FBI lab personnel – larger numbers and American Society of Crime Laboratory Directors (ASCLD) certification), have major laboratory improvements been made. From nation to nation and jurisdiction to jurisdiction within autonomous states, new laws and court-imposed standards proscribe a new set of expectations on the forensic laboratory.

For example, in the US the acceptance of testimony about novel scientific methods has experienced rapid revision. Expanded from the 1923 'Frye standard' that the scientific methodology be generally accepted, US Federal Rules of Evidence the 1993 US Supreme Court ruling in *Daubert* v. *Merrill Dow Pharmaceuticals, Inc.* (and subsequent refinements) state that courts should consider at least four aspects before admitting expert testimony. The four are: (1) whether the technique or method has been tested (i.e. reputable scientific journal articles, adequate controls used accurately, research done expressedly for litigation, blind trials, etc.); (2) whether it has a known error rate (i.e, determined through blind proficiency testing); (3) whether it has passed peer review scrutiny (without which courts should be skeptical); and (4) whether it is generally accepted by the scientific community. Laboratory accreditation facilitates many of these requirements, helping courts to make decisions about evidentiary reliability.

Internationally, the World Trade Organization (WTO) recognizes the lack of acceptance of laboratory test results as a technical barrier to international trade. Forensic science, worldwide, is benefiting from this enhanced attention to laboratory accreditation. Attention is being focused continually on the

International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC). ISO/IEC Guide 2 defines laboratory accreditation as 'formal recognition that a testing laboratory is competent to carry out specific tests or specific types of tests' and may include 'recognition of both the technical competence and impartiality of a testing laboratory.' The US Drug Enforcement Administration has provided the venue for representatives from the United Nations, Canada, United Kingdom, Japan, Germany, the Netherlands and Australia to promote quality assurance (including certification and accreditation) in forensic drug analysis, for example, through the scientific working group (SWG), acronym SWGDRUG.

In India the National Accreditation Board for Testing and Calibration Laboratories (NABL) has instituted compliance requirements with the ISO/IEC Guide 58/EN-45003 with specific reference to accreditation of forensic laboratories in ISO/IEC Guide 25/EN-45001. This includes annual inspection for quality system management requirements, maintenance of all technical parameters, and satisfactory participation in interlaboratory proficiency testing. For this purpose, the National Physical Laboratory in New Delhi has been developing Certified Reference Materials for several years. Most recently (1998) the India Ministry of Home Affairs has directed forensic science laboratories under the Bureau of Police Research and Development to prepare for NABL full accreditation, including all analyses performed. Work is continuing in the area of uniformity in the approach of forensic science laboratory inspectors, including their qualifications and training.

In Europe the International Laboratory Accreditation Cooperation (ILAC) was established in the late 1970s. More recently, recognizing the importance of laboratory accreditation, the EU has established an independent European Accreditation Advisory Board (EAAB) to work with ILAC with the aim of promoting confidence in one-stop accreditation of industry and public authorities, including forensic laboratories. Separately, the European Network of Forensic Science Institutes (ENFSI) have initiated interlaboratory projects, with member laboratories in more than 20 European countries. Specific areas of forensic expertise, such as handwriting examinations and shoeprint comparisons, have experienced difficulty standardizing quality assurance approaches specified under EN 45000 for example. In addition, standardization of proficiency testing among European countries is under development through the joint A-EUROLAB-EUROCHEM Working Group Proficiency Testing in Accreditation Procedures. In 1997 the European cooperation for Accreditation (A) merged the activities of the European Accreditation of Certification (EAC) with the European cooperation for the Accreditation of Laboratories (EAL). The United Kingdom's network of forensic science service laboratories in England and Wales have maintained accreditation for several years under the National Accreditation of Measurement and Sampling (NAMAS) M10, a service of UKAS (United Kingdom Accreditation Service), and ISO 9000 series quality standards.

Laboratory accreditation defined

Laboratory accreditation encompasses external oversight of laboratory operations, including: whether the laboratory facilities are adequate, whether the laboratory personnel have the appropriate background (expertise and experience) and opportunities for continuing education to perform assigned tasks satisfactorily, whether the laboratory has a quality control program and the degree to which this program strives to achieve excellence, how the laboratory performs on proficiency tests, how the laboratory complies with established standards as determined by laboratory inspections, and other factors that affect the reliability and accuracy of testing and reporting done by the laboratory. An example of specific requirements for urine drug testing may be used as a guide. The following paragraphs describe aspects that comprise each of these areas.

Adequate laboratory facilities

Forensic laboratories require safe, secure, and uncontaminated work areas containing the proper equipment. An unsafe forensic laboratory not only jeopardizes the health and safety of workers, but risks the compromise of evidence. Security extends from normal working hours to whenever the laboratory is closed. Visitors (including all service personnel) must be documented and escorted at all times to protect the integrity of evidence testing and chain of custody. Scientific personnel should have access restricted to only those specific areas that their work requires. After-hours security should deter and detect any unauthorized entry to the laboratory. A convenient method of restricting access and recording entry is by the use of key cards connected to a central computer system for logging. Unnecessary clutter can be unsafe, and contamination must be minimized and assessed periodically to know its potential to affect results.

Finally, a laboratory cannot meet modern analytical expectations without the proper equipment, maintained and in good working order.

Laboratory personnel

The quality of work in any laboratory rests on good personnel. The expertise and experience of laboratory personnel require opportunities for continuing education to remain current in the performance of assigned tasks. Certain accrediting organizations require academic degrees for positions such as laboratory director (doctorate) or certifying scientist (master's or bachelor's degree in chemistry, biology, or forensic science). Frequently the accreditation organizations have rigid guidelines of educational attainment for specific positions, which may underestimate the value of on-the-job experience. Not all requirements of accreditation need be imposed simultaneously. Accreditation of a laboratory must proceed slowly to allow personnel time for educational improvements and to obtain qualified personnel – to avoid disruption of laboratory services. When new accreditation agencies are forming, often current personnel who do not meet the new, more stringent requirements may be 'grand fathered', granting them a special exception for a certain time. The work skills of laboratory personnel may be brought to acceptable standards and/or improved first by probationary training periods of up to three years; in-service continuing educational programs conducted within the laboratory itself; relevant scientific seminars, conferences, symposia, and meetings; and part-time completion of degree programs.

Quality control program

A laboratory's quality control program indicates the extent to which excellence is a priority. One gauge may be the percentage of quality control samples per specimen analysis (frequently 10% in urine testing programs). Where applicable, standards and controls (from different sources) analyzed within established tolerances add confidence to identification and quantitation, as do equipment calibration records (how often, how thorough, and what steps are taken when outside of established tolerances), and maintenance records. Written standard operating procedures are necessary in order to reduce subjectivity and to provide objective analysis and interpretation of results. Because reliability decreases near the limit of detection (LOD), it is important to define what method(s) are used to measure LOD and limit of quantitation (LOQ). For example, in forensic analytical toxicology, experimentally determined LOD and LOQ use signal-to-noise ratios of 3:1 and 10:1, respectively, measured with serial dilutions in the matrix of concern. Statistically determined LOD and LOQ rely on 'quantitation' of a series of blank samples (usually at least 10), calculation of the mean (*) and standard deviation (σ), and applying the formulas: LOD = * + 3σ and LOQ = * + 10σ. Although more difficult when analyzing unique substances, written procedures that detail the approach and specific criteria for the analysis of novel analytes provide useful guidance to analysts and crucial insight to those tasked with evaluating a laboratory's performance. An approach to developing procedures for detecting novel analytes should consist, at minimum, of: searching the literature for relevant literature references; obtaining standard compounds; placement of reference compounds in a matrix similar to the specimen matrix; and analyzing with standard procedures, all before analysis of the questioned specimen. Sometimes a better approach for novel analytes would be to contact and transfer the specimen to another laboratory better equipped in its analysis. An example in forensic DNA typing is extensive quality control to assess contamination by polymerase chain reaction amplification products that could affect reliability of results and physical separation of extraction from analysis operations.

Proficiency test performance

Proficiency tests (PTs) measure laboratory performance by submitting specimens containing materials known only to the submitting agency. Open PTs are known to the laboratory to be PT samples, although the specific materials and/or their concentration are unknown. Blind PTs are submitted as any other sample from a client, so that the forensic laboratory does not recognize them as PTs. The quality of the PT program depends on the rigor of the PT challenge. For example, where cut-off concentrations are mandated by statute, such as in workplace drug testing, PT samples containing 75% and 125% of the cutoff would be more appropriate than ones containing 25% and 250% because the expectation is to distinguish concentrations at $\pm$20% around the cutoff. Similarly, PTs containing known interferences or metabolites, normally present in real specimens, represent more rigorous challenges. The frequency of PTs may vary from daily to yearly, depending on accreditation requirements and sample volume of the laboratory. Certain failures in the PT program can cause a laboratory to lose accreditation. For example, an accrediting organization may partially withdraw a

laboratory's accreditation for incorrectly quantitating a given percentage of specimens; however, laboratories can lose accreditation for reporting a single false positive result (such as reporting a drug present that is not present).

Laboratory inspections

Laboratories in an accreditation program should be inspected at least annually by an outside panel of experts knowledgeable in the work preformed by the laboratory. The inspection process should have several objective criteria such as: (1) examining recent analysis and checking for compliance with their own standard operating procedures manual (SOP); (2) does the SOP reflect established standards and current regulations? (3) where failure of the laboratory in previous proficiency testing programs has occurred, have specific steps been implemented to correct the deficiencies and are these steps adequate? The inspectors should provide a written report of the results of their inspection. Laboratory personnel should be at least familiar with how the work that they are performing compares to that done by other groups in the world. This can be accomplished by the laboratory director or members attending national and local conferences and reviewing the literature. Also, it is helpful to hold periodic staff meetings where discussions of problems with specimens are held, to review each person's tasks, and for the staff to review specific literature and discuss that literature. Such a scheme allows the staff to remain current with technology and become aware of how their role in a multistep process may impact the subsequent steps.

Compliance with established standards

Complicance is determined by the following.

1. Any other factors than affect the reliability and accuracy of testing and reporting done by the laboratory. Paperwork trail: identification of samples, custody, documentation of exceptions (had to run it again because: spilled, seal broken, quality control not right, sample volume inadequate) are the samples appropriately stored (urine frozen, DNA dried out/frozen, mold growing on things, fire debris with bacterial decomposition, time window on certain things that are known to be lost due to evaporation and/or degradation).
2. Presence of retest criteria. One example of retest criteria for urine drug testing is the presence of the drug on retesting the sample (above LOD) rather than presence above the cut-off. Criteria for if the sample fails the retest, can you explain? Example: benzoylecgonine degrades to ecgonine if the urine is basic and ecgonine is not normally detected.
3. Sample integrity checking. Are checks for adulteration routinely performed? If not, does the capability exist to conduct such tests if requested? Was the chain of custody documentation intact? Were discrepancies (broken seals; did submission form say it was dirt and it was actually paper; are signatures on custody documents? or are all samples present?) noted? Positive samples and all evidence should be saved and stored appropriately for a set period of time, in accordance with the laboratory SOP. Evidence is usually kept for one year or more and notification is normally made to the submitting organization before routine discarding of samples. Negative samples are normally discarded soon after testing.

A number of problems arise in establishing a good accreditation system for forensic laboratories. Unlike urine testing for drugs of abuse or medical samples, the work in forensic laboratories is varied and samples have a history and are frequently part of a larger body of evidence. This makes blind PT testing difficult as it would be likely that the examiner would know that a sample being submitted was a test sample rather than a case specimen. Frequently, a district attorney would not have the funds nor knowledge to submit samples to evaluate a laboratory.

Accreditation is an expensive undertaking. Smaller laboratories, associated with police units, or private laboratories may not have the resources to become accredited. Accreditation may force consolidation of the forensic testing system into state or regional laboratories. This has the advantage of concentrating resources and allowing modernization with the distinct disadvantage of loss of local control and possibly greater turn-around times. Loss of private laboratories could increase costs for private litigants and defendants. One possible solution would be for public laboratories to do private testing for a fee. This builds confidence in the public that the Government is doing a good job but opens them to criticism by unscrupulous private individuals. Participation in national round-robin tests, possibly sponsored by accreditation agencies or national standards agencies, could be an interim solution to full accreditation. Such participation would build confidence that a laboratory performed satisfactorily and make the transition to full accreditation easier or provide tiers of accreditation for forensic laboratories. A list of several accrediting organizations and their addresses is given in **Table 1**.

See also: **Quality Assurance**: QA/QC.

Table 1 Accreditation and information organizations

International Laboratory Accreditation Cooperation (ILAC)	http://www.ilac.org/	Many worldwide contacts for accreditation. Country specific contacts
International Organization for Standardization (ISO)	http://www.iso.ch/	Promulgates international laboratory accreditation standard ISO 17025. ISO 17025 is a generic standard for laboratories performing tests and calibrations.
College of American Pathologists (CAP)	College of American Pathologists 325 Waukegan Road, Northfield, IL 60093, USA http://www.cap.org/	Web site contains laboratory accreditation checklists and manual. CAP provides proficiency samples in several areas. Four separate accreditation programs: the Laboratory Accreditation Program (LAP), for all clinical laboratories, the Forensic Urine Drug Testing (FUDT) accreditation program, the Athletic Drug Testing Program (ADT), and the Reproductive Laboratory Program (RLAP), directed jointly with the American Society of Reproductive Medicine (ASRM).
American Society of Crime Laboratory Directors (ASCLD)	ASCLD, c/o NFSTC, SPJC Allstate Center 3200 34th Street South, St Petersburg, FL 33711, USA http://www.ascld.org	Certification program
American Board of Forensic Toxicology, Inc. (ABFT)	ABFT PO Box 669, Colorado Springs, CO, USA http://www.abft.org/	Certification program
U.S. Pharmacopeia (USP)	US Pharmacopeia 12601 Twinbrook Parkway, Rockville, MD 20852, USA http://www.usp.org/	Provides standard reference materials and analysis procedures for pure substances
The Substance Abuse and Mental Health Services Administration (SAMHSA)	http://www.health.org/	Workplace drug testing and guidelines. Also sets cut-off and procedures for regulated drug testing. Guidelines may be downloaded at http://www.health.org/ GDLNS-94.htm
National Institute on Drug Abuse (NIDA)	http://www.nida.nih.gov/	Provides many low cost or free publications dealing with drug use and drug testing
National Institute of Standards and Technology (NIST)	NIST 100 Bureau Drive, Gaithersburg, MD 20899-0001, USA http://www.nist.gov/ NVLAP: http://www.ts.nist.gov/nvlap	Provides standards and some testing technical documents. Manages NVLAP – National Voluntary Laboratory Accreditation Program. However, the accreditation fields are mainly in consumer products rather than areas of forensic interest
Armed Forces Institute of Pathology (AFIP)	http://www.afip.org/	Certifies and inspects Department of Defense Laboratories
Association of Official Analytical Chemists (AOAC)	AOAC International, 481 North Frederick Avenue, Suite 500, Gaithersburg, MD 20877- 2417 USA http://www.aoac.org/	Provides official analytical techniques for testing of specific materials. Laboratory accreditation in bacterial identification and detection and food analysis.

Further Reading

57 Federal Register 7139, Feb. 28, 1992.

Benjamin JT (1996) The effect of CLIA '88 and managed care on the medical lab market. Its impact on POLs, hospital labs, and reference labs. *MLO Med Lab Obs* **11**:54–58.

Roussel PL (1996) Impact of CLIA on physician office laboratories in rural Washington State. *Journal of Family Practice* **43**(3):249–254.

Tagliaro F, Smith FP, Tedeschi L *et al* (1998) Toxicological and forensic bioscience applications of advanced chromatographic and electromigration methods. In: Deyl Z, Tagliaro F and Tesarova E (eds) *Advanced Chromatographic and Electromigration Methods in BioSciences*, pp. 918–961. Amsterdam: Elsevier.

US Department of Health and Human Services (1991) *Laboratory Inspector Training Manual for the National Laboratory Certification Program*. Rockville, MD: National Institute on Drug Abuse, DHHS. Pub No (ADM) 91-1796.

ADMINISTRATION OF FORENSIC SCIENCE

Contents

An International Perspective
Organization of Laboratories

An International Perspective

J Thompson, The Forensic Science Service, London, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0400

Introduction

Events in the last two decades have changed the face of the world. There are far fewer restrictions on travel to and from countries around the world, modern communications are changing the way people do business, allowing trade to be conducted easily across time zones and when one financial market suffers the effect is felt globally.

Science generally has always been accepted as an international language and there is no reason why forensic science should be treated any differently. Whereas ordinary people around the world may be somewhat reluctant to view their role within society from a global perspective, we know that for many criminals national boundaries are no barrier. Law enforcement agencies have had to become much more active on the international scene taking part in increasing numbers of joint exercises with agencies from other countries, exchanging intelligence information and reviewing how law enforcement must change to deal effectively with international crime. Equally important is the sharing of good practise and experience as the need for universal standards becomes more recognized.

Criminal trends, social pressures and political priorities all have implications for the investigation of crime. It is now accepted internationally that forensic science has an important role to play in the investigation of crime, from the crime scene to the court room. Forensic science is committed to producing products and services which are fit for the purpose and to investing in staff to ensure that their competence and training continue to grow and keep pace with new developments in science, technology and the law. Today, that investment must transcend national boundaries.

The European Union and Europe

The last twelve years have seen dramatic changes in Europe. The Maastricht Treaty has opened the way for increased police and judicial cooperation between

the Member States of the EU; the Schengen Convention heralded the removal of border controls and greater freedom of movement on mainland Europe; and most importantly the fall of the Berlin Wall and the collapse of communism has opened Eastern Europe and new states formed from the old USSR to the West.

With these changes the European Union has experienced increased activity by organized crime from Eastern Europe within its borders. The greater freedom of movement between west and east has opened up new routes for criminals and criminal organizations trafficking in a variety of goods including drugs, firearms and people. Within the EU, the increased freedom of movement has already impacted on the investigation of organized and serious cross-border crime.

Before he was assassinated in 1992 by a Mafia bomb, Italian Prosecutor Giovanni Falcone predicted that organized crime syndicates were 'going to become the biggest beneficiaries of a free Europe'. He believed that if there were to be no frontiers for crime there should be none for justice. He called on his counterparts in Europe to organize an information network on organized crime. On 1 October 1998 Member States ratified the convention to establish Europol, an EU-wide organization responsible for information exchange and criminal intelligence analysis. Europol will provide a system for the rapid and confidential exchange of information to support multinational police action against issues such as drugs, suspected terrorists, arms shipments, asylum seekers, stolen vehicles, the illegal trade in nuclear materials and money laundering.

The EU is taking a more active role in encouraging interagency cooperation and police and judicial cooperation between Member States. The EU sees the effective tackling of serious crime, drugs and illegal immigration issues as a priority and has placed particular emphasis on practical cooperation supporting operational services provided by police, customs, other law enforcement agencies and the judiciary.

Forensic Science in Europe

In the Spring of 1992 the Gerechtelijk Laboratory in the Netherlands instigated regular meetings for forensic laboratory directors from the main public service/government forensic science laboratories in Western Europe. A year later a meeting was held in the Netherlands to discuss a European Network of Forensic Science Institutes (ENFSI). It was attended by 11 of the 13 laboratories invited. Between March 1993 and October 1995, six meetings were held and topics discussed included 'Accreditation of forensic science laboratories', 'Quality management and forensic education', 'Automation of crime laboratories' and 'International cooperation'. ENFSI was founded formally in October 1995.

Currently, ENFSI has 39 members from some 26 countries in Europe including 12 of the 15 EU Member States. The majority of members are directors of institutes who are major providers in their country of professional forensic science services to support crime investigation including the police, the judicial investigation and/or the prosecution of crime. The aim of ENFSI is to promote cooperation between its members and their laboratories through:

1. discussion of managerial questions, the effective utilization of forensic science, scientific development, and standards of practice;
2. exchange of quality assurance methods and procedures, training, scientific exchange programs and research and development;
3. cooperation with other international organizations.

To date ENFSI has taken this aim forward in a number of ways notably through its annual meeting, the work of its Board, and through technical working groups which focus on particular forensic areas. In the last two years, much has been achieved to establish a platform on which ENFSI can take forward its initiatives. In the second half of 1997, ENFSI was adopted by the European Union Police Cooperation Working Group as advisers on forensic science issues. ENFSI has also established contact with Europol, with a first meeting in September 1998. From a global perspective, in May 1997, the ENFSI Board opened up discussions internationally with other similar 'managerially' based organizations.

However, ENFSI has also been aware that the practical and technical aspects of forensic science, in particular benchwork and staff competence and training, must also be addressed on an international basis. The network has supported technical working groups since its inception and all are now addressing the issue of how forensic science can transcend international cultural and legal differences. The Strategy Plan for 1998–2003 now focuses on the business of ENFSI and in particular the promotion of good practice among forensic science practitioners.

Crime Worldwide

The problems facing law-enforcement agencies, and by association forensic science, in Europe are no different from those faced by the rest of the world.

All must be borne in mind when addressing how best to contribute to the solution.

Organized crime has always had something of an international perspective; the Colombian drug cartels, Sicilian and Italian Mafia and their American counterparts, the Chinese triads, Japanese yakuza and the Jamaican 'Yardies' are some of the more familiar organizations. Now with the changes to the former USSR and Eastern Europe we are being introduced to groups about whom we know comparatively little. Law enforcement agencies have established that these groups are active globally and are increasingly entering into business deals with each other. There is evidence to suggest that organized crime has entered into business arrangements with terrorist groups, for example arranging the supply of arms or nuclear reagents in exchange for illicit drugs, supplies of which are readily available to some terrorist groups.

Aside from organized crime, the reduction in travel restrictions and the advances in modern communications have also resulted in an increase in serious cross-border crime. Modern communications can allow individuals to commit offences such as fraud without travelling outside their own national boundaries. The ease with which individuals can now travel abroad and the high throughput experienced by some destinations, particularly the major tourist attractions, creates additional problems for an investigation.

There are individuals who specifically travel abroad so that they might indulge in practices which are considered to be abhorrent by the majority of civilized societies and which have been criminalized in their own country. There are still countries where culture may dictate an age that a girl can marry, and therefore engage in sexual relations, which is significantly younger than that which others may find acceptable. In other countries limited resources and other political priorities may prevent overstretched law-enforcement agencies from properly addressing such issues as child abuse. As a result, this has provided opportunities for Western pedophiles to continue to practice in comparative safety from prosecution. Campaigns by individuals and the United Nations have started to change the situation but, more importantly, hardening attitudes of some countries to citizens committing such crimes overseas has resulted in an unexpected response. In some countries, legislation has already been enacted which potentially allows the authorities to prosecute a citizen for certain sexual offences against children which have been committed outside their national boundaries. Changes in attitude such as these present major challenges for criminal investigations and justice systems around the world.

The forensic science community has already started to address the issues raised by these examples, what contribution it can make and what impact this will have on the administration of the science as a whole. In support of this, international organizations have begun to acknowledge the important role forensic science can play. Aside from the EU Police Cooperation Working Group, mentioned earlier, both Interpol and the United Nations have acknowledged the potential contribution of forensic science.

Interpol (the International Criminal Police Organization), which was established in 1914, is a worldwide organization (to date 178 member countries) whose aim is to promote and facilitate inquiries into criminal activities. Interpol has established working groups which focus on a number of fields of expertise within forensic science, bringing together scientists and law-enforcement officers to work in partnership.

The United Nations came into existence in 1945 and membership now stands at 185 States. The UN 'Family' includes bodies which focus on justice and human rights and the global drug problem and the involvement of organized crime. In 1996, the Secretary-General on Human Rights produced a report on 'Human rights and forensic science' further reinforcing earlier work in 1992 which acknowledged the crucial role of forensic science in identifying probable victims of human rights violations and the collection of evidence for any subsequent legal proceedings. The mission of the United Nations International Drug Control Programme (UNDCP) is to work with the nations of the world to tackle the global drug problem and its consequences. Two of the ways it intends to achieve this is by strengthening international action against drug production, trafficking and drug-related crime and by providing information, analysis and expertise on the drug issue. Both of these will benefit from the involvement of forensic science. UNDCP has also acknowledged the importance of forensic science. In 1997/98 it contributed funding towards improving standards in the provision of forensic science in the Caribbean and establishing a center of excellence.

Forensic Science Around the World

Everything mentioned so far has implications for the administration of forensic science, especially from an international perspective, and the forensic science community has already begun to respond.

The American Society of Crime Laboratory Directors (ASCLD) and Senior Managers Australian and New Zealand Forensic Laboratories (SMANZFL) are similar managerially based organizations to ENFSI.

ASCLD, a professional organization of crime laboratory managers primarily supporting forensic laboratories in North America, believes in 'excellence

through leadership in forensic science management'. ASCLD's membership also extends to Directors of forensic laboratories around the world. SMANZFL currently covers the forensic laboratories in Australia and New Zealand and aims to provide leadership in the pursuit of excellence within forensic science. All three organizations, ENFSI, SMANZFL and ASCLD, actively support technical working groups charged with taking forward the business of forensic science at the technical/scientific level and increasingly these groups are communicating with one another.

No doubt there are other similar organizations, and where they do not exist I would like to think that steps are being taken to establish them.

Many countries already have their own forensic science societies/associations whose purpose is to support the needs of forensic scientists themselves, for example, the Forensic Science Society in the United Kingdom, the Canadian Society of Forensic Sciences, the American Academy of Forensic Science. These organizations provide forensic scientists the opportunity to make contacts and exchange technical information, experience and best practice through meetings, newsletters, journals and the Internet.

In addition to these there are the international organizations which bring the global community together; Interpol holds a triennial Forensic Science Symposium attended by both forensic scientists and law-enforcement officers.

The International Association of Forensic Science (IAFS) was formed in 1966. It is unique in that it is an association in name only – it has no members, no budget and no permanent secretariat. Its stated aims and objectives are:

1. To develop the forensic sciences;
2. To assist forensic scientists and others to exchange scientific and technical information; and,
3. To organize meetings.

The IAFS achieves these primarily by means of its principal activity which is the choice of President and venue for each triennial. The triennial meeting brings together colleagues from around the world and from all fields of expertise within forensic science in its broadest sense.

The Future

This article highlights some of the main challenges facing the administration of forensic science into the twenty-first century. It also shows how forensic science is already responding to these challenges.

The global community is a reality and, although legal systems differ around the world, when an adjudication is taken which affects forensic evidence in one country it is likely to have implications for forensic science around the world. Parts of the criminal justice process are already becoming international and this, together with increasing cooperation between law enforcement agencies from different countries, means forensic scientists must address their profession from an international perspective. We are being called more frequently to provide evidence in courts outside our national boundaries and, for the foreseeable future, it is unlikely that there will be any significant harmonization of legal systems. Therefore, we must be confident that our evidence is robust enough to stand up to scrutiny in court anywhere in the world.

We must also ask how forensic science can contribute further to justice systems and to the way in which law enforcement agencies tackle international crime. In the last four years we have proved how forensic intelligence databases such as DNA, drugs and footwear can make a major contribution to the investigation of crime. The ability to exchange such information across national boundaries has now been recognized as an important tool in the fight against global crime and the forensic science community has already begun to address this problem. For example, ENFSI is responding to requests for advice on DNA and drugs intelligence databases and the exchange of information from the European Union Police Cooperation Working Group.

At managerial level, the ENFSI, ASCLD and SMANZFL summit has started to identify the core issues; sharing research; quality – a common approach through accreditation of systems and procedures to agreed international standards; exchange of best practice; and the training and competence of staff. The value of forensic science lies ultimately in the people who practice it. Forensic scientists and managers of forensic science institutes must be encouraged to communicate and cooperate on an international scale, for only with the support and involvement of both, can investment be made in the staff who are at the frontline of forensic science within criminal investigations.

Finally, I believe that many if not all the issues raised in this chapter are best tackled collectively. However, I do not agree with mandatory standards of uniformity. There is a space for a collective view on how science can contribute in terms of outcomes and benefits throughout the whole forensic process. How this is achieved will inevitably depend on national differences, although few, if any, would be able to present a sound argument against the need for quality in terms of accuracy and reliability, and a sound assessment of the strength of evidence to be adopted

by all. The future of forensic science in the twenty-first century should be determined through a united profession working collectively to determine how best they make this contribution.

See also: **Administration of Forensic Science**: Organization of Laboratories. **Deoxyribonucleic Acid**: Databanks. **Education**: An International Perspective. **History**: Forensic Sciences; Forensic Medicine. **Wildlife**. **Internet Crime**. **Stalking**.

Further Reading

Kopp I and Sprangers W (1997) *A History of ENFSI*. ENFSI publication.

Lucas DM (1998) A history of the International Association of Forensic Science. *INPALMS Newsletter* 4.

Thompson J (1998) Editorial. *Science and Justice* 38(3):141–142.

Organization of Laboratories

WJ Tilstone, NFSTC, Largo, FL, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0401

Introduction

The article discusses the way that laboratories are organized within a jurisdiction and the ways that an individual laboratory is organized. Some theoretical principles are described and compared with actual practices.

There is no common organizational system for forensic science laboratories. With very few exceptions, all are government departments or public sector agencies. Most are administratively within police departments; of those some are staffed by sworn personnel, some by civilians and some by a mixture. Others are found within civilian departments, but usually within some form of legal service agency. Some are large and can be described as 'full service,' whereas others offer a limited service on a local basis.

The spread of accreditation as a means of setting operational standards has brought little organizational uniformity, as programs are tailored to assure quality within each of the varied administrations and not to impose a standard structure.

General Issues

Accreditation standards

ISO requirements Forensic science laboratories are *testing laboratories*. Parts of their organization and management are covered by standards such as ISO Guide 25. **Table 1** describes the main organizational requirements.

Forensic specific requirements There is an international accreditation program specific to forensic science laboratories. It is that of the American Society of Crime Laboratory Directors Laboratory Accreditation Board (ASCLD/LAB). **Table 2** lists the organizational requirements of the program.

Organization Within a Jurisdiction

Administrative organization

The parent organization The accreditation standards make no administrative demands of a forensic laboratory, other than that it defines reporting relationships and takes steps to ensure that there are no conflicts of interest. The parent organizations encountered are set out in **Table 3**.

More than 200 (or about two-thirds) of the members of the American Society of Crime Laboratory Directors have addresses that identify the host agency for the laboratory as being in a police department. Over 80 have addresses that identify them as being

Table 1 Organizational requirements of testing laboratories in ISO Guide 25

The laboratory shall be organized and shall operate in such a way that it meets all the applicable requirements of the Guide.
If the laboratory is part of an organization performing activities other than testing . . . the responsibilities of all staff shall be defined in order to identify potential conflicts of interest.
Define, with the aid of organizational charts, the organization and management structure of the laboratory, its place in any parent organization, and the relations between management, technical operations, support services and the quality management system.
Specify the responsibility, authority and interrelationships of all personnel who manage, perform or verify work affecting the quality of the tests.
Appoint deputies for key personnel

Table 2 Organizational requirements from ASCLD/LAB

Organizational structure groups and personnel work in a manner to allow for efficiency taking account of the various disciplines (D)
Subordinates are accountable to one supervisor per function (I)
There are clear communication channels (D)
There is a library (I)
There are secure long- and short-term storage areas (E)
There is adequate work space for personnel (I) and equipment (D)
Design permits efficient evidence flow (I)
Access is controllable and limited (E)
Exterior exit points have adequate security control (E)
Internal areas have a lock system

Note that 'E' signifies a feature which must be present to gain accreditation; 'I' or 'D' signify features which the program deems to be Important or Desirable, respectively.

hosted in another government agency. Less than 20 are private or found in universities.

That group of over 200 contains a great variety of facilities. It ranges from the huge FBI Laboratory employing over 500 people and active in all areas, and in R & D; through the Illinois State Police system, with eight sites including a large urban facility employing about 200 staff, and a training center; to several throughout the US which employ just a few people to provide a drug and firearms service to a local police agency.

Organizational issues related to host agency location

It has been argued that there is a danger intrinsic to the location of forensic science laboratories within a police department. Namely, that they unconsciously become agents of the prosecution and lose their scientific objectivity. There is no convincing proof of this, either way. Some of the finest operational and research facilities are housed as part of a police department. The overwhelming bulk of work in forensic science is carried out on the instructions of investigators or prosecutors, irrespective of the administrative location of the service laboratory. The many private laboratories conducting DNA testing seem quite capable of working equally well, whether acting on the instructions of prosecutors, or defense.

Table 3 Parent organizations for forensic science laboratories

Police departments
Legal service agencies such as district attorney
Universities
Other government departments
Private commercial companies

This is a specific issue identified in ISO accreditation standards – potential conflicts of interest. The answer is that it is not the nature of the host agency, but the nature of the responsibility and accountability of the laboratory staff and management which counts. If it is an issue, it is not a very serious one in practice as hardly any forensic science laboratories have been set up specifically to deal with the matter of operational independence. An example is the Swedish National Laboratory of Forensic Science in Linkoping. It is housed in a University and its budget and operations are controlled by an independent board. Another example is the State Forensic Science Centre in Adelaide, Australia, which was established by legislation which required it to be free from control by police or other legal service agency.

The main problem with forensic science laboratories and location within a police agency is that of resources. Although some of the finest and best-resourced laboratories in the world are found within police agencies (the former Metropolitan Police Laboratory in London, UK; the FBI Laboratory in Washington DC; and the Forensic Science Center at Chicago, for example), there are very many examples where the police host has not been able to understand the level of resources needed to establish, maintain and develop a quality forensic science laboratory. In general, the commitment is seen by the level of reporting relationship enjoyed by the laboratory director. Agencies which take their laboratories seriously have the director reporting to someone very senior in the hierarchy.

Financial organization of forensic science laboratories One of the largest forensic science organizations in the world, the Home Office Forensic Science Service (FSS) in England, is currently organized as a government agency. This brings with it certain financial freedoms and disciplines, as it requires the FSS to operate as a business and recover its costs through the sale of services to customers.

This mode of operation was established for some services in Australia and New Zealand several years before its adoption by the FSS. It is a strange amalgamation of what most would regard as a core public service (administration of justice) being delivered by a

public entity operating to (some) private sector business standards, and has not been adopted in the US or other countries.

The rationale behind the Australian and New Zealand initiatives was that governments wished to identify the true costs of delivering public services. They thus required indirect costs such as capital depreciation, accommodation rental, leave and other staff benefits, and insurances to be quantified and included in the cost of providing the service.

The next step was to make the customer agencies responsible for meeting the costs of the service. The rationale here was that they would purchase only what they needed and could afford. The final step was to define a series of policy–purchaser–provider relationships, in which the government set policy and funded purchasers who sought the best value from available providers.

In some cases, where providers were government agencies, they were required to provide a return on capital to the government equivalent to the shareholder's return on the investment of a similar private industry. The agency was permitted to retain profit to reinvest in the business.

Some of the advantages and disadvantages of commercially operating forensic science services are set out in **Table 4**.

In theory, commercialization of services should also introduce competition. Competition in turn should mean that providers have to focus on providing only timely, necessary services which meet the standards of the customer. However, one of the customer's standards will be to get the best return for the limited budget available. The consequences of quality failure in forensic science are too dramatic for it to operate in a price-driven market. Also one of the expressed concerns about existing service structures (see above) is that forensic services provided from police agencies may suffer from unconscious bias due to their close affiliation with the prosecution. A commercial service with a strong customer service ethic will have the same problem. Finally, discontinuing laboratory investigations which are not cost effective raises concerns at the long-term loss of analytical and interpretative skills in areas such as hair examination, and therefore possibly significant evidence not being available.

Probably the two main difficulties with public sector commercial services are that those who are devoting resources to quality assurance and who are honest with their costing will suffer as they are undercut by providers – especially other public sector providers – who either bid at marginal rates which do not reflect the true operational costs, or cut corners in quality to lower costs. The second difficulty is that police investigators and the laboratory do not have the normal purchaser–provider relationship that applies to the purchase of goods or services. Agencies such as prosecutor, defense and judiciary all have a major stake in the services and can influence service delivery without having to carry any fiscal responsibility. As an example, one of the earliest commercial services was established and developed in South Australia in the period 1984–87, and still in operation there. In that time the single greatest influence on the service was the decision by the Chief Justice to impose a 6-week maximum time period from charge to first hearing at which time evidence had to be ready for scrutiny by the defense. The provider laboratory and purchaser police department were left to deal with the mandate from within existing resources.

Organization Within a Laboratory

Administrative Organization

Testing areas covered The ASCLD/LAB program defines eight subject areas or disciplines: Controlled

Table 4 Commercial operation of forensic science laboratories

Feature	*Advantage or disadvantage*
Identifies true costs	Should give basis for rational funding decisions but can distort comparisons unless applied in the same way to all laboratories.
Encourages savings	Better value may result in lower quality service as corners are cut. Avoids 'empire building'.
Customer responsiveness	Who is the customer (police, prosecutor, court)? Impact on impartiality by doing whatever the purchaser pays for.
Capital equipment depreciation	Allows better financial planning.
True service culture	Same service available to all participants in the justice system. Service provided is tailored to customer needs in areas such as timeliness. ('Timely, necessary service')

substances, toxicology, trace evidence, serology, DNA, questioned documents (QD) latent prints and firearms and toolmarks. These can be taken as defining the scope of a normal full service crime laboratory. The internal organization of a laboratory will need to allow for the handling of the work load within each discipline and the interaction between them.

Table 5 shows the disciplines and some of the more common interactions each has with one or more of the other fields. Generally, QD and firearms and toolmarks are stand alone (other than their interaction with latent prints), although gun shot residue examination may be performed in trace chemistry in some organizations, and weapons may require biological testing.

Evidence recovery Table 5 also shows the interactions between the disciplines. Effective case work management starts with effective processing of evidence recovery. Controlled substances, QD, firearms and toolmarks, latent prints and toxicology are usually self-contained in regard to evidence recovery. Cases are readily identifiable and there is no special procedure to recover evidence.

In contrast, trace, serology and DNA depend on the efficient processing of cases to recover the evidential material before the testing even starts. The physical plant and personnel organization must allow for recovery that is fast, complete and free from contamination.

Evidence management The ASCLD/LAB program places a premium on evidence integrity. The laboratory must provide secure long- and short-term storage capacities. Evidence in long-term storage must be under proper seal and evidence in short-term storage must be protected from loss, contamination and deleterious change. The main organizational aspects of meeting these standards are evidence retention and control of access to unsealed evidence.

Evidence retention is a major problem for many laboratories. There is a common tendency for law enforcement agencies to use the laboratory as a repository for evidence. Examination of the evidence store of almost any laboratory will reveal cases more then 3 years old. Many of these will have been disposed of by the court or will no longer be the subject of an active investigation. For example, the author has audited a laboratory with about 30 analysts which had an evidence store of more than 1000 ft^2 (90 m^2) to deal with the problem of retention of old cases. In contrast, a laboratory with more than 100 analysts, but a strong policy on case acceptance and evidence return, had a property room of under 600 ft^2 (55 m^2).

Examples of practices to deal with the problem include a laboratory service which delivered evidence from completed cases back to the originating police offices on a weekly basis, a routine destruction of evidence over a year old unless an active instruction is received from a senior police officer or prosecutor, and a routine screening of information on the jurisdiction's Justice Information System to identify cases disposed of.

Whatever procedures are put in place, the internal organization of the laboratory has to recognize

Table 5 The disciplines encountered in a full service crime laboratory and their interactions with each other.

Discipline	*Interacts with*	*Resulting in*
Controlled substances	Toxicology	Contamination; intelligence information to target toxicology test
	Trace	Examination of packaging materials
DNA	Latent prints	Sample preservation (effect of refrigeration which preserves biologicals but degrades latents)
Serology	DNA	Body fluid identification (potential problem where DNA analysis is separate from biological evidence screening).
	Trace	Evidence recovery and interpretation
Latent prints	Trace chemistry and serology	Latent development
	All	Latent preservation – need to protect from contamination
Trace (associative evidence)	All	Evidence recovery and contamination
Trace, DNA and serology	All	Evidence interpretation (serology and DNA are the only ones giving personal information) and analysis time (these tend to have the longest run round times)

evidence control and management as a major area of responsibility.

Internal access control The most effective way to meet the requirement for preservation of unsealed evidence in cases being worked is through control of access. The simplest way to achieve this is for each analyst to have a personal secure locker which they and only they can access. A back-up key can be retained under seal for emergency access.

Generalist or specialist Organization according to the accreditation disciplines implies that staff are specialists in each area. Not all laboratories accept this organizational assumption. Smaller laboratories require multiskilled staff to cover the range of tests they wish to offer. In other instances, such as in California, laboratories choose to develop multiskilled generalists. The argument is that it not only offers management flexibility, but the broad awareness that comes with it insures a better appreciation of evidential value.

The argument in favor of specialization is that it is not possible to gain and maintain a sufficient knowledge and practical skill base to meet the needs of operations today. Issues such as proficiency testing, the growth of specialist instrumentation, and increasing peer scientific awareness that DNA brings forth all favor specialist rather than generalist practitioners.

Centralization of services

Some laboratories serve jurisdictions which are of sufficient size in population or geographical area to be deemed to require more than one laboratory. The laboratories may all be full service (or at least offer a range of disciplines), or may be small regionals offering a limited service to the local community. There is no single answer to the optimum organization in this regard but some of the factors to be considered are shown in **Table 6**.

Discussion

A visit to 10 randomly chosen forensic science laboratories would produce 10 different ways of organizing a laboratory. About a third would be well-organized, efficient and effective facilities; about a third would be performing well but with opportunities for improvement; and the remaining one-third would be producing a service which could clearly benefit from some attention.

Consideration of the factors which can influence the way in which a laboratory is organized and set out in the tables, shows the following key areas:

- Organization to comply with external standards.
- Organization to deal with effective case and evidence processing.
- Organization to deal with the external factors influencing the laboratory such as its administrative location and the nature of the population base it serves.

External standards

Tables 1 and 2 set out some of the pertinent requirements of the ISO Guide 25 and ASCLD/LAB standards for accreditation of a laboratory. These require the organization to deal with:

Table 6 Factors in organization of laboratories within a system

Factor	*Response*
Population	Regional laboratories providing extensive service to share workload and meet needs of defined population centers.
Geographical area	Specialist regional laboratories to service local courts in areas such as blood alcohol, drugs, firearms, and latent prints. Differing needs of urban and rural communities.
Low demand but essential cases with major capital requirement	Single specialist center such as ICP-MS, or outsource testing
Technically demanding test areas where establishment costs are high	Single specialist center such as STR data basing, or outsourcing
Areas requiring rapid turn round	May be easier to establish local service than to provide special courier service for collection and reporting

- Accountability, authority and responsibility;
- Freedom from conflicts of interest.
- Organizational and lay-out structured to allow for efficiency taking account of the various disciplines (see also Table 5).
- Evidence integrity.
- The requirement for adequate work space for personnel and equipment.
- Physical security.

One of the consequences of the standards – particularly the ISO Guide 25 – is that lack of resources is not an acceptable reason for poor work. If the resources are not sufficient to permit adequate quality of work in any area, then the testing must not be performed in that area. There are many different ways to resource laboratories and no evidence that any one is superior to the others (see Table 4, for example).

Effective and efficient case processing

Effective case processing means meeting the customer's service standards. The customer assumes a zero defect service in regard to correct results from appropriate tests. After that the customer requires a timely service. The 'right' answer on the day after the suspect has left the area or the trial is over is no use. Timely results are difficult to achieve, and along with evidence management (see 'Evidence management'), require close liaison with customer agencies: police and prosecutors. In some cases efficiencies will result from centralization especially where expensive capital equipment incorporating automation is involved (for example automated immunoassays on centrifugal fast analyzers and confirmation with automated MS or MS-MS for toxicology). In other cases a small local service will be better (for example drug assay and firearms examination). Organization thus needs to be aware of:

- The range of disciplines provided and their interactions (Table 5).
- Customer service standards required.
- Technology availability and cost.
- Policies on case acceptance, prioritization and disposal.

External factors

The laboratory does not exist in isolation. It is either part of another organization with a mandate for law enforcement or laboratory services or seeks to meet a perceived demand. Demographic and geographic factors (Table 6), the nature of the parent organization (Table 3), the nature of the funding mechanism (Table 4), and the legislative and governmental norms of the jurisdiction served will all influence the way in which the laboratory is organized.

However, a review of the nature of the organization of laboratories which have been accredited leads to the conclusion that there is no optima way to organize a forensic science laboratory. However, scrutiny of the best practice examples reveals that the best laboratory:

- Reports to someone close to chief executive level in the host agency.
- Is accredited.
- Measures and responds to customer needs while maintaining scientific integrity.
- Has a significant proportion (at least 25%) of its operating budget devoted to quality assurance and training.

Laboratories which meet these benchmark standards are able to deal effectively with the requirements placed upon them by external factors.

See also: **Accreditation of Forensic Science Laboratories**. **Administration of Forensic Science**: An International Perspective.

Further Reading

Saferstein R (1998) *Criminalistics, An Introduction to Forensic Science*. 6th edn. Englewood Cliffs, NJ: Prentice-Hall.

Tilstone WJ (1991) Education, training and assessment in forensic science. *Journal of the Forensic Science Society* **31**:95–100.

Age Determination *see* **Anthropology**: Morphological Age Estimation.

Airbags *see* **Accident Investigation**: Airbag Related Injuries and Deaths.

ALCOHOL

Contents

Blood
Body Fluids
Breath
Congener Analysis
Interpretation
Post Mortem

Blood

B Hodgson, Forensic Laboratory Services, Royal Canadian Mounted Police, Ottawa, Canada

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0402

Introduction

The determination of alcohol (ethanol) in the blood of human subjects is one of the basic analytical procedures of an analytical toxicology laboratory. Alcohol is a commonly abused substance in most societies, taken deliberately for its mood altering effects.

When those effects result in legally harmful consequences to the imbiber, or others affected by the imbiber's actions, then a precise determination of the concentration of alcohol in the body is required. This specific determination allows for a proper interpretation of the degree of mood alteration. More importantly, it allows a means of assessing the degree of impairment in the ability to operate complex machines, e.g. motor vehicles.

The measurement of alcohol alteration should, in theory, be in the brain but for obvious reasons this cannot be done in live subjects nor is it necessary. The alcohol concentration in the brain has a direct correlation with the arterial blood alcohol concentration. Interestingly, the arterial blood alcohol concentration can be indirectly measured via breath. Breath measurement has become the method of choice in law enforcement of drinking-driving laws, but measurement of blood is still widely used.

Because of the high risk associated with extracting arterial blood from live subjects, venous blood samples are used for alcohol determinations. Capillary blood is another alternative and offers the advantage over venous blood of being more closely correlated to the arterial blood alcohol concentration.

Collection

Venepuncture of a cubital vein in an arm is the standard technique used clinically for extracting venous blood. This procedure offers the most convenient yet relatively safest way to draw blood from a live subject. The alcohol content of these samples will represent the systemic blood concentration of alcohol, and the measured value can be readily correlated with the behavioral effects of alcohol.

An alternative to venous blood is capillary blood obtained by finger-tip puncture. This blood, obtained from the fleshy tissue of the finger-tips, represents a transition from arterial to venous blood. As such, the alcohol concentration more closely represents the concentration in arterial blood. The major disadvantage of finger-tip blood is the very limited quantity that can be extracted. The amounts extracted are usually just sufficient to perform an initial analysis with no chance of repeat if subsequently required. Thus, venous blood has become the predominant choice for forensic purposes, since the extraction is relatively more comfortable for the subject, allows more than sufficient quantity of blood for repeated analyses, and easy correlation with behavioral effects.

In deceased subjects, a wider selection of sites for blood collection becomes available. However, the wider selection is counterbalanced by the complications in interpretation arising from the differences in alcohol concentration among the sites. As for living subjects, venous blood extracted from a vein in the arm or leg provides the optimum sample for interpretive purposes. The least desirable site is an open chest cavity that results from an autopsy examination.

Blood of uncertain origin (heart, vein or artery) may gather in the chest and mix with the interstitial fluid that bathes the organs of the chest. This mixed fluid is often scooped up and submitted for alcohol analysis as a 'blood' sample. The alcohol analysis

produces a result which presents difficulties in interpretation since the sample does not truly reflect a genuine blood origin.

In forensic work, the collection of a blood sample must observe certain precautions related to the integrity and security of the sample. This is particularly so when the result will be used in the prosecution of an offense charged against an individual, such as the operation of a motor vehicle while under the influence of alcohol. Precautions are most crucial when the offense is that of exceeding a specified statutory blood alcohol concentration (BAC) in the operation of motor vehicles.

Swabbing the injection site with an alcoholic antiseptic solution is a common practice in clinical settings. Alcohol swabbing compromises the integrity of a blood sample taken for alcohol analysis and is discouraged for forensic purposes. Unfortunately, microorganisms residing on the skin, potentially on the surfaces of apparatus used for taking the blood, or suspended in the ambient air, can contaminate the blood sample. Such microorganisms utilizing blood sugars can produce alcohol through a fermentation process. Conversely, microorganisms could also use alcohol present in the blood from drinking as an energy source. Either way, the true BAC of the person is compromised, leading to difficult interpretations.

Blood alcohol kits, such as pictured in **Fig. 1**, have been developed expressly for forensic purposes. These kits are self-contained, requiring no additional apparatus whose cleanliness and alcohol-free status may be open to question. The kits contain tubes that are sterile and contain a preservative, such as sodium fluoride with an anticoagulant. Tubes used in Canada contain sodium fluoride to produce a final concentration of 1% w/v and potassium oxalate as anticoagulant to produce a final concentration of 0.2% w/v. The preservative stabilizes the blood sample for an indefinite period of up to several months, if required. The anticoagulant prevents clotting, an undesirable feature when analyzing the blood. Although not essential, the anticoagulant nevertheless simplifies the analysis by eliminating the step of homogenizing clotted blood. This step would otherwise require the use of homogenizing apparatus, a messy and somewhat cumbersome procedure when dealing with whole blood. Refrigeration of the blood samples at approximately 4°C is recommended for prolonged

Figure 1 Blood collection kit with two vacuum blood tubes (a, b), sterile needle (c), needle holder (d), nonalcoholic swab (e), seals (f) and packaging (g, h).

storage. Experimental data have shown that blood concentrations decrease slightly over time when stored at room temperatures. The exact cause of the decrease is not certain but is thought to be either evaporation of the alcohol around the rubber stopper, or oxidation to acetaldehyde using oxygen from oxyhemoglobin (the red pigment in the blood). Refrigeration stabilizes the blood alcohol for periods of up to six months. The preservative, sodium fluoride, and the commonly used anticoagulants, do not interfere with the analytical procedures in current forensic use.

It should be noted that analysis for forensic purposes requires whole blood. Hospital and clinical labs routinely separate the fluid portion, plasma, from the blood. The plasma can be conveniently analyzed for a number of chemicals, including alcohol, using automated equipment specifically designed for that purpose. Forensic or toxicology laboratories, however, are more likely to analyze whole blood for alcohol since statutory limits for legal operation of motor vehicles are expressed in terms of whole blood, e.g. 80 mg of alcohol in 100 ml of blood. Alcohol concentrations expressed in terms of plasma have to be converted into the equivalent concentration in whole blood. Since this conversion is directly related to the water content of the blood of the individual, a degree of uncertainty is introduced at this point. The water content, or hematocrit value, varies not only among individuals but within the same individual. Hence, the equivalent whole blood concentration cannot be predicted precisely, but only within a range of values with acknowledgment to possible values outside that range.

Analysis

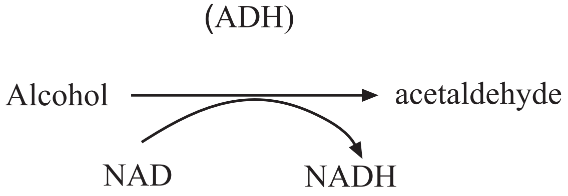
Chemical

Analysis of blood for alcohol has been performed for the better part of the twentieth century. Wet chemical methods using an oxidation–reduction reaction were developed to analyze alcohol condensate distilled from the blood. One of the more widely used methods was developed by E.M.P. Widmark, a Swedish scientist considered the classic pioneer worker in systematic studies of alcohol in the human body. This method of incubating blood samples suspended above a dichromate-acid solution at raised temperatures in enclosed flasks proved a reliable method. The procedure was modified in subsequent years, most notably in the early 1950s. This procedure, labeled the modified Widmark procedure, is still used currently for verification of alcohol content in blood standards, or as a crosscheck on other procedures.

The wet chemical methods provide satisfactory accuracy and precision, but the nature of the procedures, including corrosive chemicals, make them less than desirable to use. In addition, these procedures lack the degree of specificity for ethanol considered necessary for forensic purposes. These procedures cannot distinguish ethanol from other common alcohols such as methanol or isopropanol. Methanol is widely used as an industrial solvent, labeled as methyl hydrate, and in such products as windshield washer fluid, or gasoline line antifreeze. It is a toxic substance sometimes found in illicitly distilled alcoholic beverages. In some instances it has been added to beverages as a cheap alternative to ethanol. Isopropanol (2-propanol) is also an industrial solvent but can also be used as a rubbing alcohol in pharmaceutical preparations. Again this toxic substance is sometimes ingested resulting in poisoning in humans. Hence, the usefulness of oxidation–reduction methods in a broad forensic context is limited to samples taken from motor vehicle drivers who have consumed alcoholic beverages only. Alternative methods are required for subjects who may have ingested the other volatile substances, either deliberately or unintentionally. In post-mortem examinations, the possibility of microbial-induced deterioration of the body, or putrefaction, becomes a distinct possibility. Microorganisms not only produce ethanol but can also produce other volatile substances which can interfere with the alcohol analysis. These other volatile substances are indicative of putrefaction and can be used as confirmation for the putrefaction process. Chemical methods, such as the modified Widmark procedure, are unable to distinguish these volatile substances from ethanol and therefore are unsuitable for identifying possible putrefaction in exhibit material.

Biochemical

Enzyme oxidation is another methodology first developed in the 1950s for measuring alcohol. A number of variants have been subsequently developed, centred around the basic enzymatic reaction:



where ADH is alcohol dehydrogenase and NAD^+ is the coenzyme nicotinamide adenine dinucleotide. NAD^+ is reduced to NADH in direct proportion to the concentration of alcohol present in the sample being analyzed. The NADH can be measured by its absorption of ultraviolet (UV) radiation at 340 nm.

The ADH analysis was developed into a kit form which has found widespread use in hospital or clinical laboratories. The kit form is best suited to laboratories that only perform alcohol analysis on an infrequent basis. In these situations the use of more expensive and elaborate methods would not be feasible. In more recent years, automated analyses using batch analyzers were developed which allowed the processing of more samples.

In addition the ADH procedures were designed for measuring ethanol only, a limitation in forensic work. In the forensic field, toxicology laboratories must be capable of detecting, confirming and, if necessary, quantifying any volatile substance present that is foreign to the human body.

Instrumental

The advent of gas chromatography (GC) proved a welcome advance in forensic analysis for alcohol. Although the equipment has a high capital cost and requires high purity compressed gases, the principle of operation is relatively simple, fast, accurate and precise with a high degree of specificity.

The basic principle of GC is shown in **Fig. 2**. A sample is vaporized in an injection block and then carried through a long thin column, usually stainless steel, glass, or fibrous material, by an inert carrier gas such as nitrogen or helium. The column can either be of the packed variety, i.e. packed with an adsorbent material, or an open tube of very small internal diameter. The latter type, known as capillary columns, have internal diameters of less than 1 mm. The inner walls of capillary columns are coated with an adsorbent film.

The vaporized sample emerges from the other end of the column where the components come in contact with a detector. The volatile components of the original sample can be visually displayed either on a computer screen or on paper.

The separation and movement of the components in the blood sample are determined by the rate of carrier gas flow, the temperature of the column and the type of adsorbent material utilized. The gas flow and oven temperature can be adjusted to achieve optimum conditions for the clear separation of volatile components from one another. The adsorbent material will retain the volatile components to varying degrees depending on the degree of molecular attraction of each component with the adsorbent material. Passage of the components can take 15–20 minutes or so. Capillary columns, on the other hand, not only allow very fast passage, 5 minutes or less, but also result in sharper, cleaner separation of the components. Hence, capillary columns have become the separatory columns of choice in GC analysis.

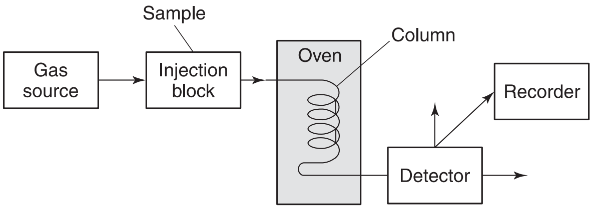


Figure 2 Schematic diagram of GC instrument.

The type of detector most commonly used is the flame ionization detector (FID). Water does not register on FID hence this detector is the one of choice for water-based samples such as blood. Although the signal produced by the detector is an analog one, the electrical impulses can be digitized. This process takes place in an analog-to-digital (A–D) converter from whence the digital impulses can be quantified. By measuring the number of impulses or counts from a known concentration of alcohol, one can compare that number with the number produced by an unknown concentration of alcohol. The concentration of the unknown can be calculated from the known since the response of the detector is directly proportional to the concentration of alcohol.

Although this quantification is a simple process, it is highly subject to variances in the volume of sample injected onto the GC column. Consequently, a more reliable technique employed by laboratories is the use of an added internal standard (ISTD). A blood sample is mixed with a solution containing another volatile compound (ISTD) such as n-propanol to produce a diluted sample. This diluted sample is easier to handle than straight blood and the volume injected onto the GC column is not critical. The ISTD must be volatile and must be clearly separable from the alcohol being measured. The ratio of counts between the ISTD and the alcohol is a constant regardless of the variation in volumes mixed or the volume of diluted sample injected onto the GC column.

Figure 3 shows a typical apparatus, known as an automatic dilutor, which will extract an aliquot of blood from the tubes featured in the background then dispense that aliquot automatically mixed with an aliquot of ISTD solution. The ratio of blood to ISTD solution is usually in the order of 1:10 to produce a sufficiently watery blood dilution to minimize the matrix effect that can affect the analytical results.

Although the diluted sample can be manually injected, this procedure is labor intensive and time consuming. Hence, manual injection has largely been replaced with automatic injector systems. These systems sit on top of a GC instrument and automatically extract small aliquots of the diluted samples and inject them onto the GC column.

The injection of diluted blood samples even with

Figure 3 Automatic diluter apparatus for mixing blood samples with ISTD solution.

automated injector systems, has the disadvantage of clogging GC columns and injection apparatus with blood proteins. These proteins shorten the life of GC columns and easily clog the small bore tubes and needles of the injection system. As a result, the technique known as headspace (HS) sampling introduced during the 1960s has become the preferred method for GC analysis of blood samples. Instead of injecting an aliquot of the diluted blood sample, a portion of the vapor sitting above the aqueous sample is injected onto the GC column. GC columns do not deteriorate as rapidly, injection systems do not clog with blood proteins, and many thousands of injections can be made on the same column.

Reliable analysis using HS-GC requires that the diluted samples be incubated in sealed vials with large headspace volumes at controlled temperatures. This incubation period allows full equilibrium of the vapor portion above the liquid sample. **Fig. 4** shows a typical HS-GC set up. The large instrument in the middle is the GC instrument itself with a control panel on the right side to set injector port, column and FID temperatures. On the left side are the controls to regulate carrier gas flow. Two sets of controls are shown since most GC instruments permit two separate columns to be housed in the same oven. To the left of the GC instrument is the headspace apparatus in which the vials are placed. This apparatus requires its own carrier gas supply which is injected into the vials to flush the headspace and carry the alcohol–ISTD vapor sample onto the GC column. The control box for the apparatus is to the left. On the right of the GC instrument is an optional integrator with printout if a computer system is not available, or cannot be justified on financial grounds.

On the extreme left of the figure is an A–D converter which relays the FID signal to a computer (not shown) which processes the signal and reports the BAC.

Ideally a GC system, especially HS-GC, will separate the common volatile substances likely to be

Figure 4 HS-GC instrument system with GC instrument (a), HS analyzer (b), A–D converter (c), and optional integrator with printout (d).

encountered in forensic work accurately and precisely. **Fig. 5** and **6** show representative separations of a mixture of four such substances together with the ISTD (n-propanol). Two different columns on different GC instruments result in different retention times (time taken for each component to move through the columns) and in a different order.

Analytical Quality Assurance

A laboratory that conducts forensic alcohol analysis on body fluids should have documented procedures in place. These procedures should identify such items as methods of analysis, quality control procedures, and performance criteria for accuracy, precision, and specificity.

A suggested protocol is outlined as follows:

I All exhibit samples should be analyzed for ethanol and other similar volatile substances using validated methods.

Validation of methods:

1. Compare the new method with an accredited method.

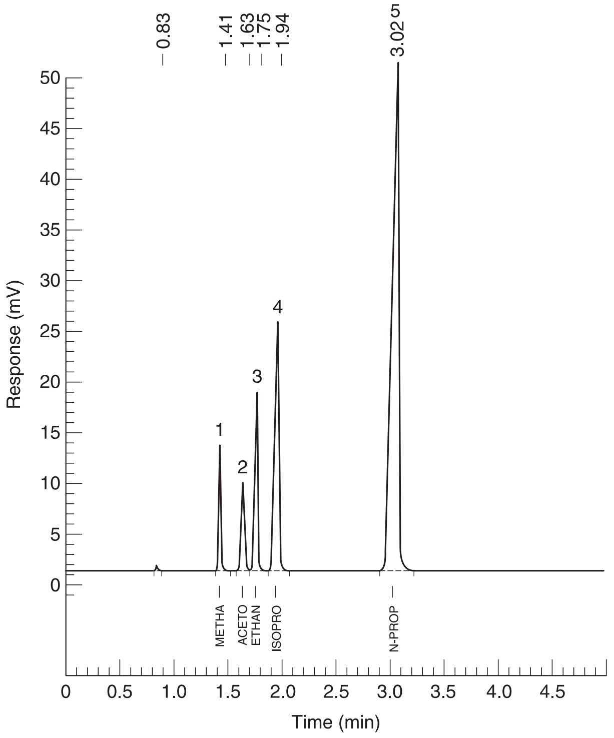


Figure 5 FID #1. Column: DB Wax × 0.32 mm × 10 m, 0.5 μm film; joined to DB-1 0.32 mm × 20 m, 5.0 μm film. 1, Methanol; 2, acetone; 3, ethanol; 4, isopropanol; 5, n-propanol (internal standard).

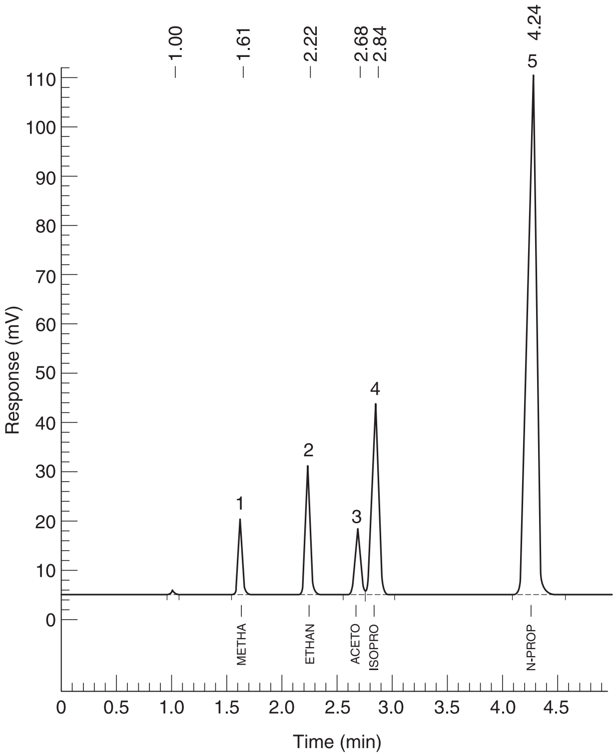


Figure 6 FID #2. Column: DB-1 0.32 mm × 30 m, 5.0 μm film. 1, Methanol; 2, ethanol; 3, acetone; 4, isopropanol; 5, n-propanol (internal standard).

2. Perform a minimum of ten replicate analysis on five different concentrations of alcohol standards specified in the range of 0–400 mg in 100 ml of solution (mg%).
   (a) Perform a statistical comparison of the data obtained to determine if there is a significant difference between the two methods; and
   (b) Perform a statistical analysis to determine the limit of detection (LOD) and the limit of quantitation (LOQ) for the method.
   (c) Determine the specificity of the method.
   (d) If no statistical difference between the new method and an accredited method is determined, and the limits of detection and quantification and the specificity are acceptable; then approve the new method for use.

II A quality control procedure of a validated method should be conducted with exhibit analysis:

1. Quality control standards could be prepared either by weight or by volume in either water purified by distillation or deionization or whole blood. (Sheep's blood is recommended, but human blood could be substituted).
2. For calibration checks, replicate analytical results

from the quality control standards with concentrations less than 100 mg% should not exceed ±3 mg% of the target value; and replicate analytical results from quality control standards with concentrations of 100 mg% or greater should not exceed ±3% of the target value.

3. When calibration check results do not meet the specified performance standards, analysts should identify and redress any nonconformance in the quality control procedure prior to conducting analysis on exhibit material.

III All exhibit material should be analyzed using at least two aliquots of the sample.

1. For volatile substance concentrations of less than 100 mg%, all replicate results should not exceed ±5 mg% of the analytical mean result. For volatile substance concentrations of 100 mg% or greater, all replicate analytical results should not exceed ±5% of the analytical mean result.
2. Any reported volatile substance in samples should be confirmed by a second method.
3. The volatile substance concentration should be reported in terms of the body fluid analyzed.
4. Ethanol, acetone, methanol and isopropanol should be reported when their concentrations exceed 3 mg%. Other volatile substances, detected and verified, should be reported when the results are of pharmacological significance.

Laboratories that conduct forensic alcohol analysis should hold available for potential court testimony all documentation pertaining to the testing procedures. All test materials should be treated as evidence with appropriate security, proper documentation, retention and storage of records and items. Laboratories engaged in forensic alcohol analysis should employ the services and advice of at least one qualified forensic scientist who specializes in the analysis of body fluids for alcohol. A qualified forensic scientist is a person who, using the appropriate combination of knowledge, skill, experience and integrity, undertakes one or more of the following tasks: analysis of evidence materials, interpretation of evidence, presentation of expert testimony.

Laboratories should participate in recognized external proficiency tests for alcohol. Such tests should be conducted on at least an annual basis. Corrective action should be taken whenever deficiencies are revealed by the proficiency test process. Analysis of exhibit material should be suspended until such deficiencies are corrected.

See also: **Alcohol**: Body Fluids; Breath; Interpretation; Post Mortem.

Further Reading

CAN-P-4C (1991) (ISO/IEC Guide 25). *General Requirements for the Accreditation of Calibration and Testing Laboratories*. Ottawa: The Standards Council of Canada.

CAN-P-1578 (1998) *Guidelines for the Accreditation of Forensic Testing Laboratories. Program Specialty Area – Forensic Testing Laboratories*. [PSA-FT]. Ottawa: The Standards Council of Canada.

Emerson VJ (1998) Alcohol analysis. In: White P (ed.) *Crime Scene to Court – The Essentials of Forensic Science*, pp. 263–288. Cambridge: The Royal Society of Chemistry.

Garriott JC (1996) *Medicolegal Aspects of Alcohol*. Lawyers & Judges: Tucson.

Jain NC (1971) Direct blood-injection method for gas chromatographic determination of alcohols and other volatile compounds. *Clinical Chemistry* **17**:82–85.

Jones AW (1996) Measuring alcohol in blood and breath for forensic purposes – a historical review. *Forensic Science Review* **8**:13–44.

Jones AW and Pounder DJ (1998) Measuring blood-alcohol concentration for clinical and forensic purposes. In: Karch SB (ed.) *Drug Abuse Handbook*, pp. 327–356. Boca Raton: CRC Press.

Machata G (1974) The advantages of automated blood alcohol determination by head space analysis. In: *Proceedings of the Sixth International Conference on Alcohol, Drugs and Traffic Safety*, Toronto, September 8–13, pp. 541–547. Toronto: Addiction Research Foundation of Ontario.

Tagliaro F, Lubli G, Ghielmi S, Franchi D and Marigo M (1992) Chromatographic methods for blood alcohol determination. *Journal of Chromatography* **580**:161–190.

Body Fluids

R C Denney, Dr. Ronald C. Denney & Co., Sevenoaks, Kent, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0403

Introduction

The most readily available drug throughout the world is alcohol (a term which throughout this article is used to mean ethyl alcohol). Despite its reputation for causing addiction, for leading to violence and disrupting family life it is a socially accepted drug in most countries of the world. The antisocial aspects relating

to its use mean that various restrictions on its availability, sale and use apply in most places where it is produced and sold. Because of its effects on the behavior and actions of individuals especially in connection with motoring offences and other more serious crimes there is frequently a requirement to ascertain what the blood, breath, saliva or urine alcohol level of a person may be or may have been at a particular time. Because of this it is necessary to know how alcohol is absorbed, distributed and eliminated in the human body, which requires access to reliable sampling and analytical procedures.

The analysis of body fluids for the presence of alcohol and other drugs is of major importance both clinically and forensically. There is no doubt that in forensic science, breath and blood have become the two most important fluids for sampling in order to carry out analyses for alcohol (ethanol) particularly in relation to motoring and other offences involving transport, assaults and even murders. However, in theory it would be possible to sample any part of the body that possesses a water content and obtain a corresponding alcohol value to those for the other tissues. In practice, it is common to carry out analyses for alcohol contents on many parts of the body during post-mortem examination and a knowledge of the interrelationship between them is useful in establishing when alcohol may have been previously consumed and the part it may have played in the individual's death.

Alcohol Absorption and Elimination

Once alcohol is taken into the body, either mixed in foods or in the more common way as in alcoholic drinks, the processes of metabolism commence. In the stomach the alcohol is subjected to three possible processes.

A small amount of the alcohol, less than 15%, is oxidized biochemically by the enzyme alcohol dehydrogenase (ADH) situated in the stomach lining. This is known as 'first pass' metabolism, but is very small compared with the more substantial metabolism of the majority of alcohol that passes into the bloodstream and that is eventually broken down by the same type of enzyme in the liver. First-pass metabolism is more significant with low levels of alcohol consumption, giving blood alcohol levels below $30\,\mathrm{mg\,dl^{-1}}$, than it is with larger intakes.

A portion of the alcohol drunk, up to about 20% for people drinking on empty stomachs, is absorbed through the walls of the stomach and passes into the bloodstream. However, this process is fairly slow and the proportion of alcohol absorbed here depends on the length of time that the alcohol stays in the stomach. This will also depend on the bulk quantity of the drink, and means that alcohol from beers and lagers will usually take longer to enter the bloodstream than that from spirits and sherry. The rate of transfer into the bloodstream depends on the ability of the ethanol molecules to penetrate the stomach lining as well as the concentration of alcohol, the related osmotic pressure and the other components, such as carbohydrates, in the drink. Alcohol taken after or with food will be held in the stomach longer and its absorption into the bloodstream will be slower than if the same amount of alcohol had been drunk on an empty stomach. Fatty, bulky foods tend to slow down alcohol absorption considerably, and under these conditions total absorption of any alcohol drunk may take more than two hours.

The majority of alcohol is absorbed rapidly through the walls of the small intestine once it has passed through the pyloric sphincter from the stomach. This rapid absorption occurs because of the extensive surface area of the villi forming the walls of the small intestine. The fastest absorption of alcohol occurs with alcoholic drinks containing about 20% alcohol by volume and with spirits mixed with carbonated diluents as these accelerate the passage from the stomach to the small intestine. Neat spirits tend to irritate the stomach and absorption is slowed by the production of a protective mucus on the absorptive surfaces of the gastrointestinal tract.

Once the alcohol has passed from the gastrointestinal tract into the bloodstream it is progressively diluted by the blood and transported around the body, passing though the liver and the heart, rapidly reaching the brain and being circulated to the other body tissues. The relatively small ethanol molecules pass very easily through the various body and cell membranes and readily penetrate to all parts of the body. At any time there is a constant transfer of alcohol from the bloodstream into any surrounding tissue. The amount of alcohol in a particular part of the body will depend on the water content of that fluid or tissue. The brain has a very high water content and is well supplied with blood vessels, so any alcohol in the bloodstream is transported rapidly to that part of the body. About 90% of the alcohol available in the bloodstream when it reaches the head readily passes into the brain tissue, at all times establishing a dynamic equilibrium between the stationary fluid content of the brain and the mobile blood movement. This state of dynamic equilibrium is maintained throughout the whole body as the consumption of alcohol continues and also as the alcohol is progressively metabolized and the body alcohol level drops.

Within the brain alcohol acts as a depressant on the central nervous system undermining normal inhibi-

tions and leading to a general deterioration of mental processes. This is due to inhibition of the action of neurotransmitters associated with brain receptors. It is unlikely that any single mechanism can explain the diversity of neurochemical actions that occur due to alcohol. It certainly has an anesthetizing effect probably due to its action on membrane processes. It is also believed that it binds to specific sites on γ-aminobutyric acid (GABA) receptors enhancing their activities as evidenced in neurochemical and behavioral studies.

With pregnant women and nursing mothers a proportion of any alcohol drunk will also pass into the fetus and the milk. This can lead to an increased likelihood of alcoholics giving birth to children with physical and mental defects.

As alcohol is continued to be drunk, a steadily increasing amount will pass into the bloodstream both from the stomach and the small intestine. Because of this the alcohol level throughout the body will not be uniform and will vary throughout the bloodstream. This state of affairs will continue while alcohol consumption continues, as there will be simultaneous elimination of alcohol from the processes of metabolism, perspiration and urination (including storage in the bladder) at the same time as alcohol is passing into the body. About 95% of the alcohol ingested is eventually metabolized in the body by enzymatic processes in the liver and about 5% is lost through breath, perspiration and urination.

The metabolic processes contributing to the elimination of alcohol from the body occur in three oxidation stages involving different enzymes in the liver. The first of these is conversion of alcohol (ethanol) into acetaldehyde (ethanal), which is then converted into acetic acid (ethanoic acid), and the final step is the oxidation of the acid to carbon dioxide and water. Although the metabolites acetaldehyde and acetic acid do not contribute to the level of intoxication of the drinker they can have unpleasant side effects. Acetaldehyde is a poison and anything greater than minor concentrations in the bloodstream will give rise to a flushed and hot complexion, breathlessness, throbbing headaches, dizziness and nausea. Under normal conditions acetaldehyde does not accumulate in the body due to the further oxidation process catalyzed by the enzyme aldehyde dehydrogenase, but this can be prevented by the use of some drugs. One of these, Antabuse (disulfiram), is specifically prescribed for patients who wish to be discouraged from drinking. The intention is that by taking the drug and continuing to drink alcohol the patients will find the whole process so unpleasant because of the headaches and other disorders that they will prefer to give up the alcohol.

It should also be emphasized that the presence of small levels of acetaldehyde and acetic acid in the bloodstream do not interfere with the analytical values for the blood, saliva or urine alcohol levels. The gas chromatographic method used for quantitative analyses involves the total separation of the alcohol from the other substances before it is individually quantified.

Dynamic Equilibrium

Once alcohol ceases to be absorbed into the bloodstream then the processes of elimination take over completely and distribution of the alcohol in the body enters into the state of ‘dynamic equilibrium’. At this stage the alcohol contents of the various tissues and fluids will bear a constant ratio to each other and this will be maintained whilst alcohol is progressively eliminated from the body, so long as additional alcohol is not drunk. Circulation of the blood maintains the ratios, effectively redistributing any alcohol around the body.

A moderate to large amount of concentrated alcohol drunk in a short period of time on an empty stomach will lead to a rapid increase in the blood alcohol level to a peak value and an uneven distribution of alcohol throughout the body. This may also lead to the creation of an ‘overshoot peak’ for a brief period (**Fig. 1**) which is above the normal shape of the theoretical absorption/elimination curve.

However, this is most commonly encountered under laboratory conditions where subjects are required to reach measurable body alcohol levels in a short space of time by rapidly drinking fairly large volumes of spirits. But laboratory conditions do not always reflect real life where it is more common for several drinks to be consumed over an extended period of time. This means that in practice the alcohol in the body may not be in a state of dynamic equilibrium

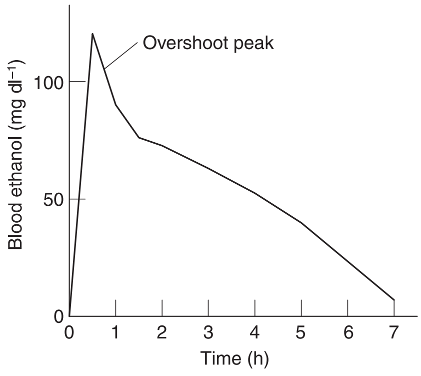


Figure 1 Alcohol absorption/elimination curve showing overshoot peak.

during the whole of the drinking session for several hours, and again will only reach this situation a short time after all alcohol drunk has passed into the bloodstream. In these cases the alcohol absorption/elimination curve will take the form of an extended arc with a central plateau level extending for an hour or more.

The curve may even have a series of minor peaks on the absorption side if there have been intervals of time between consecutive alcoholic drinks (**Fig. 2**). This will also occur if a person has been drinking moderately during lunchtime and still has alcohol in his or her body when starting to drink alcohol again in the evening – the process is known as 'topping-up'.

The sampling of body fluids and tissues during the absorption phase will lead to results being obtained which do not correspond with normal distribution ratios. For instance venous blood alcohol levels tend to be lower than the corresponding arterial blood levels during this phase. But when dynamic equilibrium has been attained it is, in theory, possible to sample any part of the body and from the measurement of the alcohol content calculate the corresponding alcohol content of some other fluid or tissue. It is this feature that enables comparable legal limits for drink-driving to be used for blood, urine and breath. However, this does not mean that medical, scientific and legal arguments do not continue about the absolute values for these interconversion ratios. As a result countries do differ in the values applied and this generally affects the comparability between the legislated blood alcohol and breath alcohol values. A typical set of values for the relative distribution of alcohol in the body is given in **Table 1**. From this it can be seen that because of its solids content whole blood has a lower alcohol content than a comparable volume of blood plasma or serum.

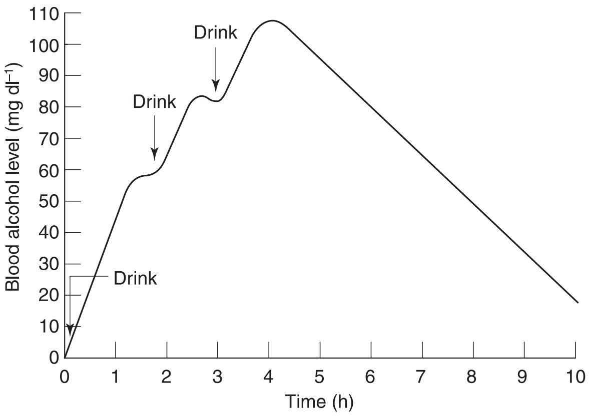


Figure 2 Alcohol absorption/elimination curve showing top-up peaks. Arrows indicate times at which alcoholic drinks were taken.

Table 1 Distribution ratios for ethanol between whole blood and other body fluids and tissues

Tissue	*Factor*
Whole blood	1.00
Cisternal spinal fluid	1.10
Urine	1.33
Saliva	1.18
Plasma or serum	1.15
Brain	0.85
Liver	0.90
Kidney	0.83
Alveolar breath*	0.000435

* This breath alcohol figure is based upon the relationship that one volume of blood contains the same amount of ethanol as do 2300 volumes of alveolar breath (the normal range of values is between 2000 and 3000 volumes).

For forensic purposes the most popular samples for alcohol analysis are undoubtedly blood and breath which are dealt with in detail elsewhere but saliva and urine are both used in different parts of the world and are of value in particular circumstances, especially as, unlike blood, they can be considered to be noninvasive. The comparability of the absorption/elimination curves for the different body fluids is shown in **Fig. 3**.

Rates of Alcohol Elimination

The amount of alcohol that is drunk even in a single measure of whisky or sherry is more than enough to overload the capacity of the enzymes in the liver. In this respect alcohol differs from other drugs, even those which are abused, in that the alcohol in alcoholic drinks is measured in grams whereas the active quantities in drugs, prescribed or illegal, are in milligrams, a ratio of 1000:1. Because there is a large excess of alcohol being pumped around the body by the heart the enzymes can only eliminate it at a constant rate. This is why the main part of the alcohol elimination curve is, essentially, a straight line. In chemical terms it follows a process of zero order kinetics. It is only at very low body alcohol levels, below about 10 mg dl^{-1}, that the elimination follows the same type of pattern as normal drug elimination with an exponential curve typical of first order kinetics.

In practice what this means is that for most of the alcohol elimination phase it is possible to predict what the alcohol level of a particular body fluid will be at a specific time if the rate of alcohol elimination has been established for part of the elimination phase of the curve. It is normally accepted that the average rate of alcohol elimination corresponds to a reduction

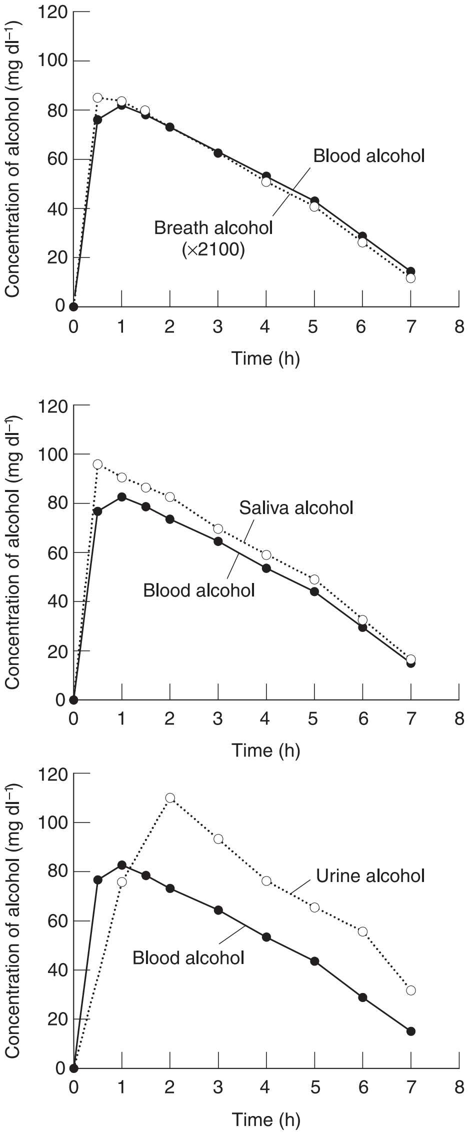


Figure 3 Comparability of alcohol levels between body fluids. (Taken from Karch SB (ed.) (1998) *Drug Abuse Handbook*, p. 340. Boca Raton: CRC Press.)

of 15 mg dl^{-1} of blood (or 20 mg dl^{-1} of urine) every hour. The equivalent figure for breath is 6.5 μg dl^{-1} of breath. This corresponds approximately to the amount of alcohol in a single drink, i.e. a single measure of spirits, a small glass of wine or about 250 ml of normal strength beer or lager. But this is an average for *all* people including occasional drinkers. For people who regularly drink alcohol the average figure is nearer 18 mg dl^{-1} for blood (24 mg dl^{-1} for urine). But the potential range is very wide extending from low values of about 9 mg dl^{-1} to very high elimination rates sometimes in excess of 25 mg dl^{-1} with some problem drinkers and alcoholics.

Alcohol in Saliva

At present this is the least used of the readily available body fluids. Its main limitation being that it cannot be provided in any substantial volume. With modern gas chromatographic analysis equipment the small volume constitutes no problem for quantitative evidential analysis. But splitting the sample between containers and ensuring that loss of alcohol does not readily occur during storage does mean that there are potential sources of error. For these reasons the main studies on saliva have been as a potential noninvasive screening sample. It is now possible to take a swab from the mouth using a cotton wool padded stick which is then inserted into a recess in a plastic plate containing ADH, plus other reagents, to produce a color reaction which is used to establish the corresponding blood alcohol level. However, few legislators appear to be prepared to include saliva in the list of body fluids that can be sampled for forensic purposes, although it does have equal potential as a readily available sample for the screening of drugs.

Alcohol in Urine

In contrast with saliva, urine is one of the fluids that has been approved for sampling purposes almost as much as blood, probably because of its ease of collection and the volume that is likely to be available. However, for quantitative alcohol analyses urine sampling has to be carefully carried out and its relationship with the corresponding blood alcohol level considered with caution.

Technically speaking urine is a 'dead liquid' when it is provided. Before that stage when it is stored in the bladder it is not in dynamic equilibrium with the other body fluids except at the time that it is actually discharged from the kidney into the ureter to travel to the bladder. This means that the bladder is receiving a constantly changing concentration of alcohol depending on whether alcohol absorption is still continuing, at the same time as alcohol metabolism, or if absorption has ceased and only elimination is taking place. The situation can be even more confusing if the urine sample is collected from a bladder which has been filled during part of both the absorption and elimination alcohol phases. So the alcohol concentration in the bladder can be increasing or decreasing and this will also depend on when the bladder was last emptied.

In many respects urine is unreliable as a medium for quantitative analyses of alcohol especially as the peak urine alcohol level occurs about 30 minutes after the peak blood alcohol level. Because of these problems,

where urine is accepted as a forensic sample for quantitative analysis, it is necessary to have subjects empty their bladders completely before waiting for 30–60 minutes before another sample for analysis is provided. However, even that later sample will be one which has varied in composition during the interval of time which has elapsed before being supplied.

Alcohol Concentrations and Calculations

Because human beings differ widely in height, weight, sex, body fat content and health, as well as in many other respects, the identical amount of alcohol drunk by different people will produce different body alcohol levels, irrespective of which fluid is sampled. In fact the same amount of alcohol may be contained in different volumes of a range of alcoholic drinks. So one person might drink 15 g of alcohol contained in about half a litre of lager whereas another person will drink the same amount in roughly a double measure (50 ml) of gin. In the first case, because of the bulk of the liquid, it will take longer for the alcohol to be absorbed into the bloodstream than it would in the case of the alcohol from the gin where it is contained in a much smaller volume. Although this again will depend on the speed of drinking. The number of variables is therefore high and accounts to some extent for people drinking the same amount of alcohol producing different blood, breath and urine alcohol levels and exhibiting different degrees of intoxication.

It has been well established that, in general, compared with men, women only require two-thirds of the amount of alcohol to produce the same body alcohol levels. This is because the average woman's body contains a higher proportion of body fat and a lower proportion of water than does the average man's body. So there is less water in the female body to dilute the alcohol.

Various attempts have been made over the years to relate alcohol levels to the amount that has been drunk. The pioneer work in this field was carried out in the 1930s by E. M. P. Widmark. His studies led to the development of the following equation:

$$c = \frac{100 \times a}{w \times r}$$

where c is the anticipated blood alcohol concentration ($\mathrm{mg\,dl^{-1}}$), but can be the corresponding breath, saliva or urine alcohol level by use of the appropriate conversion factor; g is the number of grams of alcohol consumed; w is the person's bodyweight in kilograms; r is the Widmark factor, for men this is 0.68 and for women 0.55.

The difference between the Widmark factors for the two sexes takes into consideration the relative amounts of body fat. However, these are only average values and in recent years various attempts have been made to produce individual factors for each person, where necessary, which more completely take into consideration the person's height, weight and age. These necessitate calculating the individual's amount of body water after determining the proportion of body fat. For obese people the modified factor can be 10–20% lower than the Widmark factor.

The above equation only provides a theoretical value for the possible alcohol level produced from alcohol consumption, it does not take into consideration alcohol that might have been lost due to metabolism. But this is another aspect that has to be considered if attempts are to be made to determine whether or not a person's claimed alcohol consumption corresponds to the values obtained on analysis. For instance, using the basic equation given above, a man weighing 75 kg drinking 50 g of alcohol in three half litres of lager will produce a theoretical blood alcohol level of 98 $\mathrm{mg\,dl^{-1}}$. The same amount of alcohol drunk by a 55 kg woman will produce a corresponding blood alcohol level of 165 $\mathrm{mg\,dl^{-1}}$.

However, these figures will bear no relationship to the actual blood alcohol levels, say, four hours later and some assumption has to be made with respect to the amount of alcohol eliminated. If in both cases it is assumed that there is an average rate of loss corresponding to a reduction of 15 $\mathrm{mg\,dl^{-1}}$ of blood every hour in the blood alcohol level then, in its simplest form, this would mean that the man could expect to have a blood alcohol level of 38 $\mathrm{mg\,dl^{-1}}$ after that period of time and the woman would expect to be at 105 $\mathrm{mg\,dl^{-1}}$.

This places a greater responsibility on women to take even more care about how much alcohol they drink than it does on men. Although there is no safe body alcohol level for either sex when it comes to driving motor vehicles or handling dangerous machinery. Alcohol at any level in the body can lead to impairment of abilities and judgment.

Calculations on blood, breath and urine alcohol levels are used extensively in court defenses where the defendants claim that they drank alcohol after they had been driving and were below the legal limit at the time of the alleged offence. Such calculations have to be considered with considerable caution as the large number of variables, including the quality of the evidence provided, means that these can only be considered to be a guide and indication at the best, although in practice with reliable data the variables do tend to cancel each other out. However, even apparently satisfactory values should not be accepted in the absence of additional data and evidence.

Recovering from Alcohol

Unfortunately there are very few substances that will genuinely assist in increasing the metabolic rate or speeding the elimination on alcohol in any way. Although the caffeine in large quantities of strong coffee or tea may actually make an intoxicated person feel a little better due to the stimulating effect of the caffeine it does nothing to reduce the actual body alcohol level or to improve that person's ability to drive a motor vehicle or to be responsible for their actions. The tea or coffee may, however, serve to dilute any alcohol still remaining in the stomach and hence slow down the rate of the alcohol absorption and reduce the peak body alcohol level attained.

However, in cases in which it is essential to reduce the alcohol level, as with casualties in hospital who may be suffering from head injuries, it is usual to drip feed a concentrated solution of fructose into the veins. This has the effect of accelerating alcohol elimination and enables hospital staff to establish at an early stage if the person's actions and demeanor are due to intoxication or the brain injury. However, the quantities of fructose used for this purpose would be enough to make a person sick if taken orally. This concept has led to the marketing of so called 'soberers' which are claimed to assist drinkers to recover from their excesses. These are usually tablets consisting of mixtures of either fructose and vitamin C or of kaolin and charcoal, but there is no indication that in this form or these small quantities they have any effect either on alcohol absorption or elimination. Only time and the body's natural processes will clear the alcohol from the system.

Alcohol is the most studied of all the drugs because of its ubiquity and its propensity for being abused. Other drugs tend to target particular sites in the body, but its solubility and small molecular size readily enable alcohol to reach all parts of the body. Forensic scientists continue to investigate the properties of alcohol in order to obtain an even clearer understanding of its pharmacology.

See also: **Accident Investigation**: Motor Vehicle. **Alcohol**: Blood; Breath; Interpretation; Congener Analysis. **Toxicology**: Equine Drug Testing; Overview; Methods of Analysis – Ante Mortem; Methods of Analysis – Post Mortem; Interpretation of Results; Inhalants.

Further Reading

Denney RC (1997) *None For The Road*. Crayford: Shaw.

Emerson VJ (1998) Alcohol analysis. In: White P (ed.) *Crime Scene to Court – The Essentials of Forensic Science*. Cambridge: The Royal Society of Chemistry.

Fields R (1992) *Drugs and Alcohol in Perspective*. Washington: Wm. C. Brown.

Fishman R (1988) Alcohol and alcoholism. In: Snyder SH and Lader MH (eds) *The Encyclopedia of Psychoactive Drugs*. London, Toronto and New York: Burke.

Jones AW and Pounder DJ (1998) Measuring blood-alcohol concentration for clinical and forensic purposes. In: Karch SB (ed.) *Drug Abuse Handbook*, pp. 327–355. Boca Raton: CRC Press.

Julien RM (1995) *A Primer Of Drug Action*, 7th edn. pp. 101–123. New York: W. H. Freeman.

Knight B (1991) *Simpson's Forensic Medicine*, 10th edn. pp. 273–280. London: Edward Arnold.

Martin CS (1998) Measuring acute alcohol impairment. In: Karch SB (ed.) *Drug Abuse Handbook*, pp. 307–326. Boca Raton: CRC Press.

McLay WDS (1990) *Clinical Forensic Medicine*, pp. 91–102. London & New York: Pinter.

Rand HP and Dale MM (1991) *Pharmacology*, 2nd edn, pp. 885–893. Edinburgh: Churchill Livingstone.

Walls HJ and Brownlie AR (1985) *Drink, Drugs & Driving*, 2nd edn. London: Sweet & Maxwell.

Wills S (1997) *Drugs of Abuse*, pp. 164–176. London: The Pharmaceutical Press.

Breath

R G Gullberg, Washington State Patrol, Seattle, WA, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0404

Introduction

During the past 50 years in North America, breath has emerged as the biological specimen of choice for the forensic analysis of alcohol ('alcohol' will refer to ethyl alcohol) in routine drunk driving cases. Amid some controversy, forensic breath alcohol analysis has only recently become more widespread in Europe where blood and urine alcohol programs have long-been established. Despite many legal challenges regarding perceived deficiencies in breath alcohol measurement, expanding forensic application continues. There are many reasons for preferring breath over other biological specimens including: (1) less invasive sampling, (2) rapid analysis and reporting of results, (3) expanding legal support and acceptance, (4) impairment research relying predominantly on breath alcohol data, (5) minimal training of operators, (6) easily adapted to field testing environments

and (7) health risks associated with the collection and preservation of biological fluids. Forensic breath alcohol measurement, therefore, merits special consideration and discussion.

This article presents the historical, biological, analytical, administrative, legal and future considerations of forensic breath alcohol analysis. Similar to other measurements in forensic toxicology, forensic breath alcohol analysis is concerned with the measurement and interpretation of alcohol collected from a human biological system and applied within a legal context generally defining its acceptability and interpretation. This unavoidable blending of science and law adds a unique dimension and challenge to the forensic scientist who must measure and interpret breath alcohol results.

Historical Perspective

The seminal work of E.M.P. Widmark in Sweden during the 1920s regarding human alcohol physiology and analytical measurement led to widespread development of blood alcohol programs in Europe and North America. The published work of Liljestrand and Linde in 1930 cohesively established much of the physiological theory underlying breath alcohol measurement, prompting further investigation and instrumental development. Practical methods for measuring breath alcohol were later enhanced during the 1930s when Harger developed an instrument named the Drunkometer. Jones provides an excellent overview for the historical development of all aspects of forensic alcohol analysis.

The earlier development of blood alcohol methodology resulted in the unfortunate notion that breath alcohol would be a surrogate technique and interpreted as a blood alcohol equivalent. Much debate (even to this day) has occurred in both the scientific and legal communities regarding blood and breath alcohol comparisons. Numerous legal challenges and a myriad of diverse court opinions have resulted. In an effort to remedy the confusion, legislation in both North America and Europe has been enacted which prohibits specific alcohol concentrations for both breath and blood separately. These legal measures have appropriately distinguished breath as a separate and reliably measured biological specimen worthy of forensic application and interpretation.

Biological Principles

Following the typical route of oral ingestion, alcohol will distribute itself by simple diffusion throughout all of the body water. Blood will transport alcohol through the pulmonary circulation to the lungs where partitioning by simple diffusion occurs with the associated alveolar and bronchial air as governed by Henry's Law. The blood/air partition coefficient (Ostwald coefficient) is approximately $\lambda_{ba} = 1750$ at 37°C. Relative to other respiratory gases, alcohol will by far be the most abundant when consumed to forensically significant concentrations. Moreover, various pathological conditions will not generally preclude the accurate measurement of breath alcohol concentrations. The predictable physiological characteristics of alcohol in breath have facilitated its ability to be reliably sampled and quantified following a sustained exhalation.

Proper sampling is a critical consideration in forensic breath alcohol measurement that contributes significantly to its interpretation. Breath is highly heterogenous with regard to alcohol due largely to its high solubility in body fluids. Because of significant airway interaction on exhalation, samples will generally have lower alcohol concentrations compared to that of the alveoli. **Fig. 1** illustrates continuous time sampling curves for both a human and simulator sample source. Simulator devices containing known concentrations of alcohol and water at controlled temperatures are used to provide vapor alcohol samples for instrument calibration and testing purposes. Greater variability is clearly observed in the human breath sample. Moreover, any unusual pre-exhalation breathing patterns (e.g. hyper- or hypoventilation) will result in even greater sampling variability. This sampling (biological) component of breath alcohol measurement generally contributes over 80% of the total method variability as shown by summing independent variances. Since a breath sample must be obtained from a conscious and cooperative individual (often difficult when intoxicated), the combination of instrumental sampling features along with careful interaction between operator and subject is critical

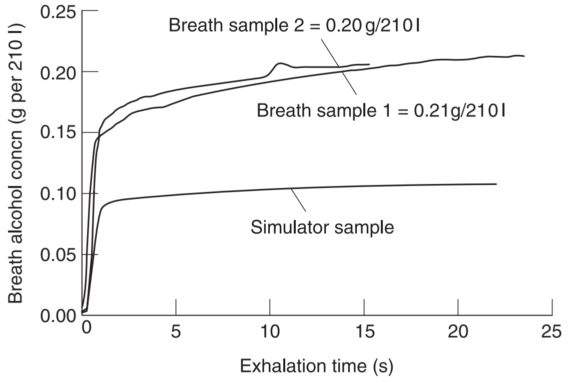


Figure 1 Continuous time exhalation profiles from a human subject and a simulator device.

to obtain consistent and properly interpretable samples that merit forensic application. As with measurement in respiratory spirometry, care and attention is necessary for adequate sampling.

Analytical Methods

Many recent advances in analytical technology have found relevant applications in breath alcohol instrumentation. Most of the commercially available instruments today employ one of the following technologies.

1. Wet chemistry: alcohol in the breath sample is oxidized in a solution containing an oxidizing reagent (e.g. potassium dichromate) resulting in a color change measured optically in accordance with Beer's Law.
2. Infrared: alcohol in the breath sample is quantified by absorbing infrared energy filtered to specified wavelengths in accordance with Beer's Law.
3. Electrochemical: alcohol in the breath sample is oxidized on a chemically active surface yielding electrons and resulting in a measurable increase of electrical conductance.
4. Gas chromatography: alcohol in the breath sample is passed through a treated column resulting in separation and eventual quantitation by, for example, a flame-ionization detector.
5. Dual technology: two technologies might be employed to improve the specificity for ethyl alcohol in breath samples. Infrared and electrochemical methods, for example, might both be used for quantitation and identification.

While employing one or more of the above methodologies, manufacturers add several features to enhance the forensic uniqueness of their instruments. In an effort to obtain deep lung end-expiratory samples, hardware and software features incorporate various parameters including: minimum flow rate, minimum volume, minimum exhalation time, alcohol curve slope determination, etc. Sample chamber purging and signal zeroing features are added to ensure reference levels for subsequent measurements. Various internal standards are added to ensure the proper function of optical, hardware and software features. Some infrared instruments employ filters for two or more wavelengths while attempting to improve specificity for ethyl alcohol. Automatic sampling and analysis of external standards (e.g. simulator or gas standard devices) are also added features to most modern computerized instruments. Various optical pathlengths and signal processing features are also added to improve signal-to-noise ratios, accuracy and precision. Computerization has probably advanced modern instruments the most while greatly enhancing their analytical and data collection capabilities. **Table 1** lists several commonly used breath alcohol instruments in North America and Europe along with their analytical method. This list is by no means exhaustive. Generally, manufacturers are willing to customize their instruments to accommodate the unique needs of local jurisdictions.

Screening and evidential purposes

Analytical methods also vary depending on whether the intended purpose is screening or evidential. The notion of screening or evidential is largely determined by relevant rules and regulations in each jurisdiction. The same device may be employed for screening in one jurisdiction and evidential in another. Screening devices are frequently handheld and employ electrochemical technology. They are generally more operator dependent and may lack some features (e.g. sampling control, external and internal standards, etc.) because of their limited purpose. Their advantage is portability, ease of operation, and rapid analysis. Results from evidential instruments are generally used in court to prove elements of an alleged criminal offense and thereby designed to address various forensic issues likely to arise. Evidential instruments will generally be computer controlled with features including external and internal standards, sampling control, specific protocol control, automatic purging and zeroing, data collection, error detection, quality control features and printout results. Assistance for local jurisdictions in the selection and use of breath alcohol equipment is provided by national organizations in the United States (National Highway Traffic Safety Administration), in Canada (Alcohol Test Committee of the Canadian Society of Forensic Sciences) and in Europe (International Organization of Legal Metrology, OIML). These organizations have developed recommended analytical standards along with approved product lists.

Table 1 Several of the most commonly employed forensic breath alcohol instruments along with their analytical method

Instrument	*Analytical method*
Alco-Analyzer 2100	Gas chromatography
Alcolmeter	Fuel cell
Alcomonitor CC	Fuel cell
Alco-Sensor IV	Fuel cell
Alcotest 7110	Infrared/fuel cell
Breathalyzer	Wet chemistry
BAC Datamaster	Infrared
Camec	Infrared
Intoximeter ECIR	Infrared/fuel cell
Intoxilyzer 5000	Infrared

Reporting results

Although different instruments offer a variety of display and printing formats, local administrative rules will generally govern those employed. Although some screening devices have a colored light system to report results, evidential instruments generally display and print results with their units truncated to two or three decimal places. Printout documents received from the instrument at the time of analysis are generally the best legal record for court presentation, avoiding further risk of error associated with transcribing, data transmission, processing and subsequent printing. The format and information content of printout documents require careful consideration because of their persuasive influence in court. Although determined largely by local regulations and issues, the information on printout documents should include: date, instrument serial number, operator name, subject name, subject date of birth, simulator standard batch number, simulator temperature verification, presampling observation time period, blank test results, breath test results, simulator standard results, associated times and units for all results.

Interpreting results

Similar to other considerations, the interpretation of results often depends on local administrative rules. Although most jurisdictions directly report the breath alcohol concentration (e.g. grams per 210 liters), some remain who must convert the value to a blood alcohol equivalent according to some assumed conversion factor. Appropriate conversion factors have generally been determined experimentally by comparing simultaneously collected blood and breath samples as illustrated in **Fig. 2**. The limitations of this approach due to analytical and biological variability should be appreciated. Although a large body of paired blood and breath alcohol data exists showing reasonable agreement, this conversion practice is forensically unsound. Additional uncertainty and unnecessary debate are introduced with this indirect approach to measurement interpretation.

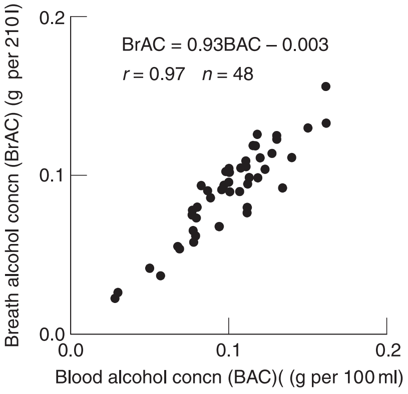


Figure 2 Scatterplot along with linear regression equation comparing essentially simultaneously collected within-subject blood and breath alcohol samples.

Computerized instrumentation

Computerized breath alcohol instrumentation offers several advantages. The measurement process can be monitored and controlled to insure conformance with quality control standards. Algorithms controlling breath sampling, signal processing, measurement protocol, error detection, reporting format, etc. can also be incorporated. In addition, data collection via modem and host computer can be accomplished. Subsequent data analysis can provide useful insight for program evaluation along with demographic and law enforcement features of drunk-driving enforcement.

Forensic considerations

Many legal challenges regarding breath alcohol have prompted manufacturers to incorporate hardware and software features addressing these matters. Minimum breath sampling parameters, for example, improve the precision among replicate breath alcohol analyses. Protocols requiring replicate breath samples along with internal and external standard and blank analyses are encoded in software requiring the operator to follow the specified test sequence. Hardware and software features allow the instrument to monitor for interfering volatile organic compounds or radio frequency interference (RFI) with the subsequent aborting of a test when detected. Other error detection capabilities are also included that monitor various instrument and test criteria. Basic design features including separation of circuitry, power line filtering, breath tube heating, continuous conductivity metal casing, stable power supply and cooled infrared detectors also minimize noise and potential interference. Dual technology and/or multiple infrared filters are often incorporated to improve specificity for ethyl alcohol. Potential interference from the multitude of endogenous volatile organic compounds is generally mitigated because of the trace levels (well below levels of detection) of these materials in normal human breath. Protocol details such as precluding the subject from observing test results until printed can also be added to minimize bias. Instruments can even be programmed to compute and report confidence intervals for mean breath alcohol results. Modern computerized instrumentation has generally enhanced the integrity of breath alcohol measurement by addressing many forensic concerns.

Analytical Protocols

Forensic breath alcohol results should be viewed as the product of an integrated measurement system including hardware, software, standards, operators, protocol, training, etc. **Fig. 3** illustrates various components of this total measurement system which all contribute to measurement reliability. Moreover, the forensic integrity and interpretation of results depends equally on competent instrumentation and testing protocol. Often, jurisdictions allocate significant time and resources in selecting the best possible instrument for their purposes but fail to give equal attention to their evidential testing protocols. **Fig. 4** illustrates the balanced consideration of both instrumentation and protocols to optimize forensic integrity and quality control.

The forensic context must be thoroughly appreciated to properly develop a breath-testing protocol that is fit-for-purpose. Other clinical measurement contexts are often inadequate models for forensic purposes since program details and interpretation of results are rarely challenged in a legal context. The profound implications of breath alcohol results should motivate a sound analytical protocol for each subject tested including: replicate breath sample analyses, replicate agreement criteria determined as a function of concentration, simulator or gas standard control measurements, blank tests, internal standard tests, presampling observation time period, instrumental and procedural error detection features and printout results. Moreover, the forthright acknowledgement when individual test results or procedures are inadequate and the full disclosure of all testing results and program policies are critical to forensic integrity. Significant legal challenge can, in some cases, motivate necessary change and improve overall quality control. Generally, forensic breath alcohol programs develop in correspondence to the intensity of legal challenge. Forensic officials would be wise to listen to the voice of their critics as well as their advocates since even the best of programs can be improved.

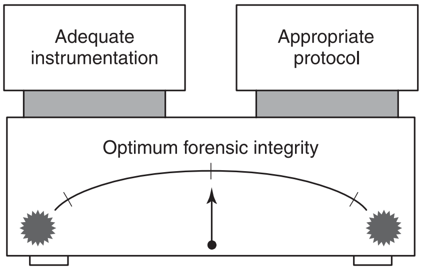


Figure 4 Forensic measurement quality control and confidence is a balance between both adequate instrumentation and appropriate measurement protocol.

The instrumental and protocol features necessary to achieve an acceptable level of quality control is largely determined within each jurisdiction depending on: statutory language, administrative rules, legal case law, common defense challenges, political and administrative considerations, training requirements, program costs and funding sources, etc. Several features that will generally enhance forensic integrity include: carefully drafted administrative rules, formally approved and tested instrumentation, formally trained and approved personnel, formally approved training outlines, formally approved instrument software versions, formally approved test protocol and

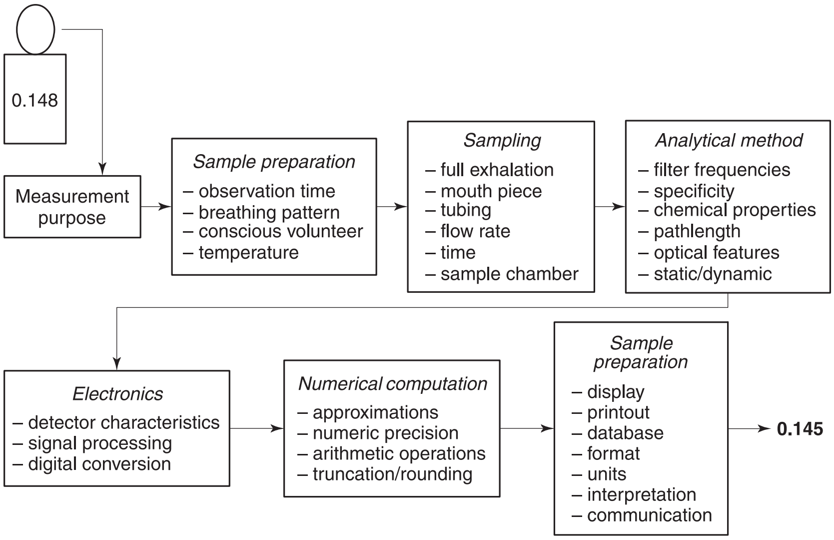


Figure 3 Breath alcohol measurement algorithm showing critical elements that comprise the total measurement system.

standards, etc. These elements comprise the 'information matrix' available to the forensic scientist who interprets and communicates test results in court. The critically important interpretation and clear communication of breath alcohol results is the final stage of the total measurement process. Without careful consideration and appreciation, more 'noise' may be introduced into breath alcohol results at this final stage of communication than during all previous stages of sample collection and analysis.

Program Administration

Forensic breath alcohol programs in North America can be administered in part or in whole at either local, state or federal levels. European programs, on the other hand, are generally administered at the federal level. Administrative rules defining program structure and policy have varying degrees of detail. These rules must be carefully constructed in view of the law and case law history since they generally define the legal admissibility and are often the focus of legal challenge. Technical detail should be minimized. Development of adequate administrative rules requires the collaborative effort of law enforcement, scientific personnel, prosecuting authorities, defense interests, court personnel, funding agencies, etc.

Two important agencies in the United States that have assisted jurisdictions in program development and administration include the National Highway Traffic Safety Administration (NHTSA) and The National Safety Council Committee on Alcohol and Other Drugs (CAOD). NHTSA has developed instrument testing protocols as well as an approved products list identifying instruments complying with published standards. The CAOD has published guidelines, policy statements and recommendations regarding many aspects of forensic breath alcohol programs.

The increased application of data collection systems has resulted from the capabilities of modern computerized instruments. Many interested agencies find applications for these data to evaluate program features including: law enforcement effort, instrument performance, demographics of arrested subjects, trends in program performance, legislative issues, defense challenges and court admissibility and interpretation. The current intrigue with computer technology is likely to increase the demand for forensic breath alcohol data collection and analysis.

Since most forensic breath-testing programs are operated by governmental agencies, program costs are an important consideration when competing for limited resources. The costs of instrumentation, personnel, training, maintenance and protocol features must all be balanced against quality control and overall forensic objectives. Many jurisdictions have cleverly supplemented their funding with penalty fees collected from the convicted drunk driver.

Legal Context

The relevance and analytical detail of forensic breath alcohol measurement is more fully appreciated within the broader adversarial legal context. To a large extent, breath alcohol testing programs, including instrumental and protocol features, are the product of continuing adversarial challenge. Program administrators often find it easier to address challenges by changing some feature of the instrument or protocol rather than continually debating the merits of the challenge. The forensic context adds a dynamic element continually challenging program structure and policy.

Typical statutory language

The trend in recent years has been to express the drunk-driving offense as prohibiting separate specified blood and breath alcohol concentrations. These '*per se*' expressions of separately prohibited concentrations have been attempts to simplify the analytical interpretation and avoid unwarranted and time consuming argument. *Per se* language, however, has not been without difficulty. A more focused challenge on all aspects of the breath alcohol test has resulted. A *per se* law provides little else to challenge.

The scientific and legal justification for these statutory measures has been due, in part, to several large epidemiological studies revealing breath alcohol as a suitable surrogate measure for driving impairment. The Uniform Vehicle Code illustrates statutory language closely followed by many jurisdictions in the United States: 'A person shall not drive or be in actual physical control of any vehicle while: 1. The alcohol concentration in such person's blood or breath is 0.08 or more based on the definition of blood and breath units in §11-903(a) (5)'. Moreover, many jurisdictions define different offenses and allow varying penalties at different breath alcohol concentrations. Implied Consent statutes are an important adjunct allowing law enforcement personnel to obtain breath samples more readily from arrested subjects. Federal funding in the United States has also been an important motivation for states to enact specified legislation in several drunk-driving related areas. This increased legal and public attention on breath alcohol analysis through more complex laws, enhanced penalties, and funding enticements are important considerations for program development and forensic application.

Case law history

Compared to other nations, the United States (with its litigious propensity) probably has the largest body of case law regarding forensic breath alcohol analysis. The case law in each jurisdiction will generally be unique because of statutory language and case law history. Major federal cases including *US* v. *Frye* (1925) and *Daubert* v. *Merrel Dow* (1993) have influenced constitutional issues regarding the scientific acceptability and legal admissibility of breath alcohol results. The case law history is an important consideration influencing program development, forensic admissibility and interpretation in specific legal contexts.

Common legal challenges and responses

The serious consequences of a drunk-driving conviction have evoked vigorous and exhaustive defense challenge to forensic breath alcohol analysis. Although some challenges are unique to specific jurisdictions, many are more universal. Several of these legal challenges along with typical prosecution responses include the following.

1. Biological issues: the individual may have an elevated body temperature, deficient respiratory capacity, abnormal breathing pattern, abnormal hematocrit, 'mouth alcohol', etc. The typical response is to argue that the law does not distinguish between these issues, therefore, they are not a defense. These matters should go to the weight of the evidence and not their admissibility. Some biological issues (e.g. abnormal breathing pattern and 'mouth alcohol') are controlled by instrumental and protocol considerations.
2. Interfering substances: the individual may have volatile organic compounds in their breath from either endogenous or exogenous sources. Research evidence tends to minimize the potential for bias resulting from exposure to these substances when proper forensic protocols are followed. Instruments are designed, generally, to preclude bias from specific interfering substances.
3. Results near '*per se*' level: results just exceeding *per se* levels may be in error with the possibility that the 'true' value is below. The uncertainty in breath alcohol measurement must be acknowledged. Appropriate testing protocols can allow for estimates of systematic and random error and determination of confidence intervals.
4. Measurement performed subsequent to time of driving: Results at the time of analysis are argued to be higher than at the time of driving where laws prohibit specified breath alcohol levels 'at the time of driving'. Many jurisdictions have enacted laws prohibiting specific breath alcohol levels within some specified time of driving (e.g. within two hours). Moreover, experimental evidence suggests that analytical results subsequent to driving will be equal or less.
5. Alcohol in the breath does not produce impairment: breath alcohol has long been a recognized index of impairment with most court appeals upholding the constitutionality of breath-specific *per se* laws.
6. Analytical deficiencies: alleged deficiencies in instrument performance have somehow biased the defendant's results. This is common where database and other records are retained. Generally, these issues go to the weight and not admissibility of the evidence. Appropriate analytical protocols with a clear presentation of the basis for confident results in court will generally overcome these issues.

Many legal challenges are overcome by the prepared, clear and honest expert testimony of the forensic scientist. Expert forensic testimony that explains clearly the basis for confidence emerging from an analytically sound program is very persuasive to the courts.

Future Forensic Considerations

Forensic breath alcohol programs today are encountering increased public attention along with added legal challenge and complexity. Protocols that minimize operator involvement and incorporate error detection capabilities should enhance confidence when facing legal scrutiny. As legal *per se* concentrations decrease, levels of detection and quantitation must be evaluated for specific instruments and testing protocols. All program elements must be continually evaluated to insure forensic integrity.

The forensic application of breath alcohol analyses must continually monitor and thoughtfully consider the relevance of advancing technology. Many aspects of 'intelligent measurement' will find important and useful applications in breath alcohol instrumentation. The flexibility of computerized instruments allows for: (1) monitoring simulator standard values and duplicate test agreement; (2) adjusting for known bias; (3) computing confidence intervals; (4) requesting additional samples depending on specified criteria, etc. Many features of 'intelligent measurement' can enhance the forensic integrity of breath alcohol results.

Improving analytical specificity for ethyl alcohol will continue to be an important motivation in the

future. This issue will likely be addressed from both instrumental and legal perspectives. Multiple infrared filters and dual technology illustrate instrumental responses to the issue while both statutory revision and case law will contribute to the legal perspective. Moreover, further research will be necessary to determine the biological and kinetic details following exposure to volatile organic compounds and the likelihood of even having measurable levels in the breath of drunk drivers.

The analytical process should become as direct as possible while attempting to measure exactly what is defined in the law. Statutes that prohibit specific breath alcohol concentrations simplify and improve the analytical process and interpretation. On the other hand, laws attempting to improve the estimation of blood alcohol equivalent concentrations by measuring and correcting for breath temperatures appear only to complicate and add considerable uncertainty to the process.

The need may arise to formulate a legal definition of 'breath' for forensic purposes. A person may be above or below the *per se* limit depending on length and technique of exhalation. The inherent variability of alcohol concentration in a single exhalation has prompted legal debate in this regard. Instrumental sampling features may have to be combined with a legal definition to clarify the issue.

The ultimate purpose of forensic breath alcohol measurement is to present reliable information that will facilitate an informed decision consistent with the relevant law. Future efforts must be directed toward improving this communication process. The use of visual aids, selected analogies, simplifying terms, and honest clear presentation will help to avoid confusion with technical material. The informative transformation of measurement results into clear and relevant terms must be a continual motivation for the forensic scientist who communicates and interprets breath alcohol results.

See also: **Alcohol**: Blood. **Evidence**: Statistical Interpretation of Evidence/Bayesian Analysis. **Toxicology**: Overview; Interpretation of Results.

Further Reading

Andreasson R and Jones AW (1996) The life and work of Erik M. P. Widmark. *American Journal of Forensic Medicine and Pathology*. **17**:177–190.

Borkenstein RF, Crowther RF, Shumate RP, Ziel WB and Zylman R (1974) The role of the drinking driver in traffic accidents (The Grand Rapids Study). *Blutalkohol* **11**:1–131.

Dubowski KM (1970) Measurement of ethyl alcohol in breath. In: Sunderman FW and Sunderman Jr FW (eds) *Laboratory Diagnosis of Diseases Caused By Toxic Agents*, pp. 316–342. St Louis, MO: Warren H. Green.

Dubowski KM (1991) *The Technology of Breath-Alcohol Analysis*. Rockville, MA: National Institute on Alcohol Abuse and Alcoholism, US Department of Health and Human Services.

Harger RN (1931) Science news. *Science* **73**:189–210.

Henry W (1803) Experiments on the quantity of gases absorbed by water, at different temperatures, and under different pressures. *Philosophical Transactions of the Royal Society of London* Part **I**:29–43.

Jones AW (1991) Top ten defence challenges among drinking drivers in Sweden. *Medicine, Science and the Law* **31**:229–238.

Jones AW (1996) Measuring alcohol in blood and breath for forensic purposes – a historical review. *Forensic Science Review* **8**:1–31.

Liljestrand G and Linde P (1930) Uber die Ausscheidung des Alkohols mit der Exspirationsluft. *Scandinavian Archives of Physiology* **60**:273.

National Committee on Uniform Traffic Laws and Ordinances (1998) *Uniform Vehicle Code*. Evanston, IL: Traffic Institute, Northwestern University.

Taylor JK (1987) *Quality Assurance of Chemical Measurements*. Chelsea, MI: Lewis Publishers.

Widmark EMP (1981) *Principles and Applications of Medicolegal Alcohol Determination*. Davis, CA: Biomedical Publications.

Congener Analysis

W Bonte, Institute of Forensic Medicine, Heinrich Heine University of Duesseldorf, Duesseldorf, Germany

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0408

Introduction

Alcoholic beverages contain besides ethanol up to 800 compounds which are responsible for their specific taste and smell. The most important compounds are the congener alcohols. These are methanol and some higher aliphatic alcohols known as fusel alcohols. In different classes of beverages, such as beer or whisky, there are characteristic qualitative and quantitative differences in the content of these alcohols. Modern analytical methods allow their detection in the blood of drunken drivers. When pharmacokinetic and metabolic peculiarities are taken into account, congener findings can be useful for the verification or refutation of allegations concerning the alcohol intake.

Congener Content of Alcoholic Beverages

Most of the flavor compounds of alcoholic beverages are present in extremely low concentrations (**Table 1**), but some can be found in sufficient quantities to allow detection of them in the blood of consumers. These are the congener alcohols: methanol, propan-1-ol, butan-1-ol, butan-2-ol, isobutanol, 2-methylbutan-1-ol and 3-methylbutan-1-ol. Methanol is a breakdown product of the vegetable pectin. The higher aliphatic alcohols are byproducts of the fermentation process and have their origin in amino acids whose amino groups are needed for the nitrogen supply of yeast. As the higher aliphatic alcohols were originally extracted from fusel oils (i.e. the third phase of the distillation procedure) they are also called fusel alcohols.

There are characteristic differences in the congener content of alcoholic beverages (**Table 2**). Fermented beverages such as beer, wine, rice wine etc. contain only small amounts of methanol, propan-1-ol, isobutanol and of the isoamyl alcohols. Distilled spirits can be nearly free of congener alcohols (e.g. gin, vodka), or contain the same congener alcohols as fermented beverages, but in different, sometimes very high, quantities. Other spirits, such as fruit brandies are characterized by extremely high concentrations of methanol or contain butan-2-ol as a specific component. Usually the concentrations of the congener alcohols are distributed according to Gauss's law within the same class of beverages.

Analytical Separation of Congener Alcohols

The sensitivity of the conventional head-space gas chromatographic system is sufficient for blood ethanol determinations down to $0.01\,g\,l^{-1}$ but for the detection of congener alcohols in body fluids the limits of detection must be reduced to $0.01\,mg\,l^{-1}$. This can only partly be managed by optimization of the analytical procedure. Additionally an improvement of the sample preparation is necessary. One possibility is to increase the sample volume from 0.2 to 1 ml and to add 1 g of potassium carbonate or sodium sulfate to the sample which causes an increased vapor pressure. This effect is more successful when handling aqueous solutions than whole blood, therefore it is necessary to subject the samples to homogenization by ultrasound and ultrafiltration. The final volume of the aqueous phase containing the congener alcohols is about 0.5 ml. As the long-chain alcohols are partly or completely bound to glururonic acid the sample should be incubated with β-glucuronidase to split the coupled compounds before salt is added.

The disadvantage of this procedure is that 2 ml of substrate is necessary to analyze either the free or the coupled alcohols, and 4 ml for both when two analyses are required. When the samples are less than this, the problem can be solved by miniaturization. Since only a small fraction of the head-space vapour is injected onto the column, 2 ml vials with only 0.1 ml of blood extract can be taken instead of the usual 20 ml vials. One needs only an aluminum cylinder in the shape of the orginal vial and a central bore for mounting the 2 ml vial.

The increased sensitivity of this method on the other hand has the effect that not only the congener alcohols appear in the chromatogram but also their metabolites and also some esters and endogenous compounds as well. This makes qualitative identification problematic. The problem can be solved by the introduction of a two-column technique with two different columns and a split before the columns. Because of the complicated extraction of the samples an internal standard (pentan-2-ol) should be added.

Another problem is that it takes nearly 60 min for the complete elution of all alcohols from both columns which causes broadening and diminishing of the last peaks resulting in a marked deterioration of sensitivity. This effect can be overcome by replacing the packed columns with capillaries. But the combination of capillary columns with the head-space procedure is delicate. Because of the small load capacity of these columns only short injection times up to 3 s are tolerable. Even splitless injection does not allow the volumes necessary for trace analyses to be injected. Increasing the flow rate of the carrier gas helium has no advantage because the retention times become too short for sufficient separation of all compounds. A good solution is to include a high-pressure valve in the system which is able to create a sample pressure of 450 kPa without influencing the flow rate.

The best results can be obtained with a cryofocusing technique. Behind the injector is installed a neutral 1 m capillary precolumn with CP-Sil-5 linked over a 1:1 split to the two separating fused silica capillaries with CP-Sil-19 and CP-Wax-52, respectively. The precolumn is inserted in a Teflon tube through which refrigerated gaseous nitrogen is passed in the opposite direction to the carrier gas flow. The end of the cool trap opens into the oven near the

Table 1 Flavour compounds of alcoholic beverages

	Concn. ($mg\,l^{-1}$)
Congener alcohols	<5 000.0
Aldehydes, carbon acids	<2.0
Higher esters	<0.001

Table 2 Congener content (mg l^{-1}) of alcoholic beverages

Beverage Class	n	*Methanol*	*Propan-1-ol*	*Butan-1-ol*	*Butan-2-ol*	*Isobutanol*	*2-Methylbutan-1-ol*	*3-Methylbutan-1-ol*
Beer, b.f.[a]	545	7±3[b]	13±3	0	0	13±4	13±3	52±12
Beer, t.f.[c]	102	1–19[d]	10–73	0	0	9–90	10–49	37–124
Red wine	282	104±30	30±7	0–7	0–1	47±15	33±6	137±26
White wine	460	29±16	31±8	0–8	0	56±18	27±6	112±25
Rosé wine	16	12–41	15–30	0	0	24–99	4–73	34–314
Champagne	53	16±8	31±5	0–9	0	53±15	30±7	126±28
French cognac	25	273±97	184±23	0–1	0–6	385±56	125±33	764±86
Wine brandy	25	272±76	130±22	0–4	1–18	252±38	77–186	482–961
Williams	18	3783±983	1382±686	211±148	609±479	210±71	65±21	307±131
Cherry brandy	19	2202±355	2218±1227	1–17	121±30	151±36	60±15	307±80
Plum brandy	16	3697±881	1039±688	53±33	64±56	230±96	62±18	341±38
Slivovitz	15	1513-4037	326–1506	7–145	26–187	58–301	28–72	139–386
Caribb. rum	27	6–74	34–3633	0–1	0–126	8–455	0–219	0–788
Blended Scotch	50	112±19	171±29	0	0	263±36	59±12	239±44
Scotch malt	23	159±9	131±16	0	0	376±34	118±10	486±40
Irish whiskey	6	6–110	170–205	0	0	284–413	72–89	288–338
US whiskey	22	196–328	50–193	0	0	388±99	271±53	1059±188
Canadian whisky	8	70–90	13–82	0	0	20–50	26–55	102–197
Gin	14	12–1359	0–885	0–1	0	0–7	0–19	0–53
Vodka	27	1–170	0–16	0	0	0	0	0
Dutch Genever	8	38–95	0–74	0	0	0–106	0–67	0–138

[a] b.f.= bottom-fermented; [b] mean ± SD; [c] t.f.=top-fermented; [d] range.

injector. Refrigeration is managed by passing gaseous nitrogen through a cooling coil inside a Dewar container with liquid nitrogen. The valve can be controlled to open at the beginning of the pressurizing phase and close at the end of the injection phase. With a pressurizing time of 2.2 min and an injection time of 0.1 min the temperature at the outlet of the cool trap is about −60°C. Sample temperature is 80°C with a thermostatting time of 32 min. The method works with a temperature program: 45°C isothermal for 11 minutes, 10°C min^{-1} to 100°C, 5 minutes isothermal, then 30°C min^{-1} to 120°C and 7.8 min isothermal. All fusel alcohols are eluted from both columns within 30 min. For the detection two FIDs (250°C) are necessary. An example of a resulting chromatogram is shown in **Fig. 1**.

Toxicity of Congener Alcohols

Richardson's law, formulated in the middle of the nineteenth century, implies that the toxicity of alcohols increases with the length of the carbon chain. Side chains as well as hydroxyl groups in a secondary position on the other hand seem to diminish toxicity. These observations provoked some speculation as to whether the acute or the postacute effects of alcoholic beverages may at least partly be due to the congener alcohols. Acute toxicity of various beverages has been little studied but the results support the conclusion that compounds other than ethanol have little effect. Reports of less serious intoxication after beer than after brandy or vodka in similar quantities seem to be mainly explained by absorption effects.

On the other hand marked hangover symptoms were observed in the post-ethanol phase after consumption of congener-rich beverages. The fact that, during this hangover phase, neither ethanol nor other aliphatic alcohols could be traced supports the hypothesis that the pharmacologically active compounds in this regard are not the alcohols themselves, but their metabolites, particularly aldehydes.

Recent studies confirm the hypothesis that the accumulation of methanol by exogenous supply, endogenous sources or both, caused by a competitive inhibition with ethanol leads during the postalcoholic phase to a production of the pharmacologically highly active metabolite formaldehyde, when the inhibition of alcohol dehydrogenase by ethanol has ceased. This would explain the heavier hangover symptoms after the consumption of congener-rich beverages, which usually also contain higher amounts of methanol, and also the well-known phenomenon that hangover symptoms can be diminished by new intake of alcoholic beverages.

Following this hypothesis that apart from acetaldehyde formaldehyde in particular could possibly be responsible for severe damage after heavy and chronic drinking, the intensive studies about the mutagenic, carcinogenic and teratogenic effects of alcoholic beverages could be seen in quite a new light. The high incidence of increased risk of cancer of the upper digestive tract after long-term consumption of home-distilled alcoholic beverages observed in specific geographic regions across the world has mainly been related to acetaldehyde; but possibly it can also be attributed to formaldehyde. In blood and in cell cultures formaldehyde showed strong mutagenic and carcinogenic properties which were not shown by the corresponding alcohols and higher aldehydes. The same circumstances might be valid in cases of fetal alcohol syndrome or alcoholembryopathia which according to some authors are caused by the primary metabolite of ethanol, acetaldehyde.

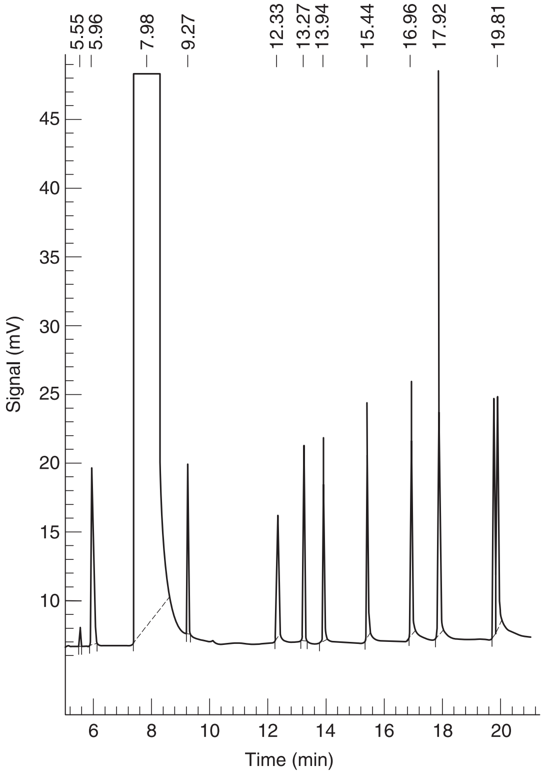


Figure 1 Gas chromatogram of a calibration standard. Capillary column CP-Sil-19, 50 m, FID. From left to right: methanol, ethanol, propan-2-ol, propan-1-ol, ethylmethylketone, butan-2-ol, isobutanol, butan-1-ol, pentan-2-ol (inner standard), 3-methyl-butan-1-ol, 2-methyl-butan-1-ol. Concentrations: methanol 4 mg l^{-1}, ethanol 1 mg l^{-1}, pentan-2-ol 2 mg l^{-1}, others 1 mg l^{-1}

Pharmacokinetics and Metabolism

It has been demonstrated that congener alcohols are resistant to storage alterations in blood samples and

can be distinguished from putrefactive products. It therefore seemed to be of forensic importance to look at the congener content of blood samples of drunken drivers. Analytical data may possibly be helpful for the evaluation of allegations concerning the alcohol intake. The question was whether there is a close correlation between the consumed amount of a congener alcohol, and the resulting blood level. Initially it was shown that there were characteristic relations, but these could not be described according to Widmark's laws. This observation made extensive experiments necessary to elucidate the pharmacokinetics and metabolism of the congener alcohols.

Information about the pharmacokinetics of the congener alcohols is mostly derived from drinking experiments with healthy volunteers. After drinking, the alcohols were rapidly absorbed through the gastrointestinal mucosa and distributed throughout the body. Methanol, propan-1-ol, butan-2-ol and ethanol are only soluble in body water. Isobutanol and the isoamyl alcohols have a much greater distribution volume because they are also soluble in lipids (**Fig. 2**). The fusel alcohols (not methanol) were partly or completely conjugated with glucuronic acid, this effect being more significant the longer the carbon chain of the alcohol (**Fig. 3**) and accompanied by an increase in the water solubility. However, only 5–10% of the consumed congener alcohols were excreted in free or conjugated form in the urine. This means that their elimination is mainly due to metabolism.

Normally the metabolism of congener alcohols as of ethanol is nearly exclusively catalyzed by liver ADH. Because of this competition inhibitory influences can be expected. There are two controversial effects, which complicate this intricate matter. On the one hand the affinity of ADH to aliphatic alcohols increases with the carbon-chain length; the oxidation of a short molecule will be (partly) inhibited by a higher alcohol. On the other hand the high concentration of ethanol in comparison to the very low congener-alcohol levels is of importance.

Both effects could be demonstrated. In general, the longer the carbon chain the faster the congener alcohol is eliminated from the blood. But the elimination speed is also influenced by the actual ethanol concentration. Very low concentrations (below $0.5\,g\,kg^{-1}$) have nearly no inhibitory effect, which results in a rapid elimination of all fusel alcohols. In consequence these alcohols can only be detected when significant ethanol levels are present. However, only such levels are of forensic interest. With higher ethanol levels the elimination kinetics of all fusel alcohols follows a first-order reaction in contrast to the zero-order reaction of ethanol detected by Widmark. Examples are given in **Fig. 4**.

The elimination kinetics of methanol is completely different. During the first two or three hours after cessation of drinking slight increases can be seen followed by a more or less constant plateau phase. Elimination only starts when the blood ethanol concentration decreased to about $0.4\,g\,kg^{-1}$. The reason for this is the twofold handicap resulting from the lower ADH affinity and the great concentration differences.

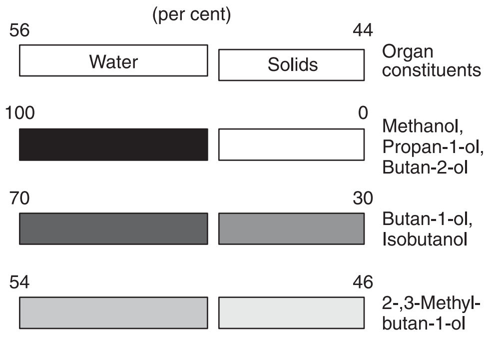


Figure 2 Solubility of congener alcohols.

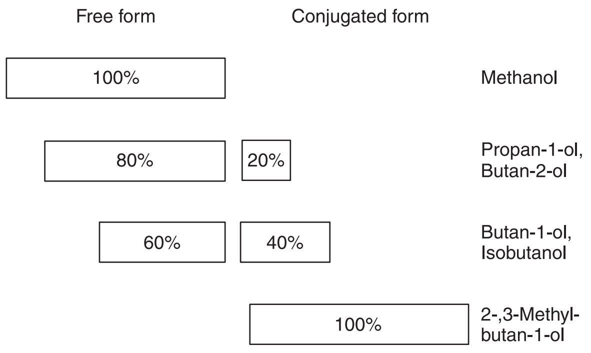


Figure 3 Conjugation rate of congener alcohols with glucuronic acid.

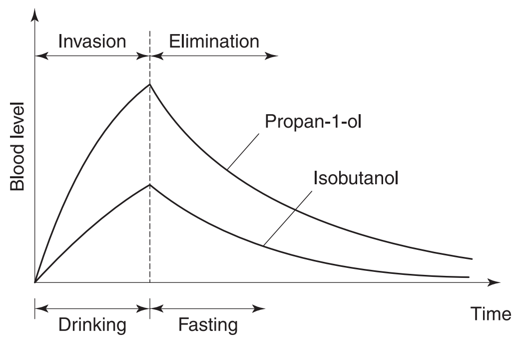


Figure 4 Invasion–elimination kinetics of propan-1-ol and isobutanol, schematized.

The special pharmacokinetics of methanol conditioned by competitive inhibition causes accumulation of methanol once continuous consumption of alcoholic beverages is practiced as long as the blood ethanol level remains above $0.4\,g\,kg^{-1}$. In its most extreme form this behavior will be found among chronic alcoholics. The question is whether high blood methanol concentrations indicate chronic abuse. Analysis of blood samples from intoxicated chronic alcoholics were compared with those from drinking experiments with healthy volunteers. Most volunteers had methanol levels of less than $3\,mg\,l^{-1}$, some had levels up to $5\,mg\,l^{-1}$ and very few had higher levels (**Fig. 5**). The highest was $6.2\,mg\,l^{-1}$. It is remarkable that more than 60% of the chronic alcoholics had values above $5\,mg\,l^{-1}$ and nearly 40% exceeded $10\,mg\,l^{-1}$ (Fig. 5). The highest value was more than $120\,mg\,l^{-1}$.

This observation confirmed the hypothesis, that blood methanol concentrations above $10\,mg\,l^{-1}$ indicate chronic abuse. Even levels above $5\,mg\,l^{-1}$ can be found nearly exclusively in alcoholics and should cause suspicion. However, the drinking behavior and the kind of beverages have a significant influence on the resulting methanol level.

Until recently it was not possible to define clear pharmacokinetic formulas for the invasion and elimination of all congener alcohols. Instead, correlation formulas for the calculation of some blood congener concentrations from drinking amounts were derived from drinking experiments (**Table 3**). The formulas are fixed on certain times after the end of drinking. For times in between intrapolation is necessary. Careful extrapolation is possible. Drinking time should not be less than one or more than three hours.

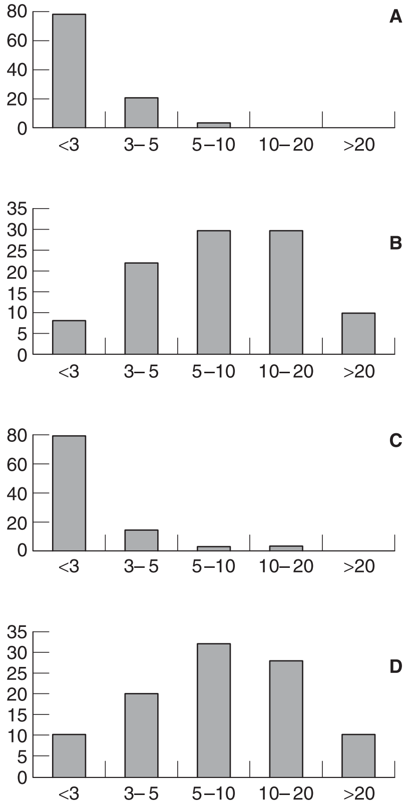


Figure 5 Distribution (percent) of methanol concentrations ($mg\,l^{-1}$) in blood samples of healthy subjects (**A**, *n*=720), chronic alcoholics (**B**, *n*=110), drivers with blood-alcohol concentrations below $2\,g\,kg^{-1}$ (**C,** *n*=1000) and drivers with blood-alcohol concentrations above $2.5\,g\,kg^{-1}$(**D**, *n*=1000).

Forensic Significance of Congener Analysis

The reason to start congener analyses in the late 1970s was to contribute to an everyday problem

Table 3 Calculation of expected blood-congener concentrations from drinking amounts $\pm$ standard deviation 30 (C_{30}), 90 (C_{90}) and 150 (C_{150}) min after drinking end

Calculation of C_0 according to Widmark's formula:

$$C_o = \frac{\text{Amount of each congener consumed [mg]}}{\text{Body weight [kg]} * r_A}$$

Calculation of blood-congener concentrations ($mg\,l^{-1}$):

Methanol	*Propan-1-ol*
$r_{A\,min} = 0.6$; $r_{A\,med} = 0.7$; $r_{A\,max} = 0.8$	$r_{A\,min} = 0.6$; $r_{A\,med} = 0.7$; $r_{A\,max} = 0.8$
$C_{30} = 0.79 \times C_0 + 0.01 \pm 0.58$	$C_{30} = 0.72 \times C_0 \pm 0.05$
$C_{90} = 0.89 \times C_0 + 0.08 \pm 0.44$	$C_{90} = 0.59 \times C_0 + 0.01 \pm 0.07$
$C_{150} = 0.95 \times C_0 + 0.16 \pm 0.28$	$C_{150} = 0.48 \times C_0 + 0.01 \pm 0.12$
Isobutanol	*3-Methylbutan-1-ol*
$r_{A\,min} = 1.1$; $r_{A\,med} = 1.3$; $r_{A\,max} = 1.5$	$r_{A\,min} = 1.6$; $r_{A\,med} = 2.0$; $r_{A\,max} = 2.4$
$C_{30} = 0.56 \times C_0 + 0.03 \pm 0.11$	$C_{30} = 0.32 \times C_0 \pm 0.05$
$C_{90} = 0.40 \times C_0 + 0.03 \pm 0.09$	$C_{90} = 0.15 \times C_0 \pm 0.04$
$C_{150} = 0.30 \times C_0 \pm 0.04$	$C_{150} = 0.07 \times C_0 \pm 0.06$

which seemed to be a specific German problem because of legal peculiarities. It deals with the so-called hit-and-run delinquency. In Germany many drunken drivers prefer to leave the scene of an accident even when the other party needs help. The fear of disqualification from driving is greater than that of punishment for hit-and-run driving. When caught one or two hours later with alcohol in their blood many of these drivers claim to have been sober at the time of the accident and only started to drink after the accident, because of excitement.

This behavior is so characteristic that a special expression was born in the German language ('Nachtrunk') which has no counterpart in English; perhaps 'postconsumption'. There was no method for the verification or refutation of such allegations. This is exactly the field where congener analysis can contribute to a clear decision. To demonstrate how such analytical results can be used for checking the truth of a claim it seems appropriate to present typical examples of actual forensic cases.

Example 1 A driver obviously did not perceive that a traffic light turned to red and crashed into a halting car. He drove back, made a U-turn and left the scene. The police were called, got the car number from the other driver involved and obtained the name and address of the car owner from the registration files. The address was checked at regular intervals but nobody opened after ringing. There was no damaged car in the street. Only four hours after the accident was the door opened. The car owner pretended to be totally surprised. He smelled of alcohol but alleged that he had drunk a lot of whiskey after midnight in his apartment. When questioned about the accident he claimed to have been at home all the time. His car must have been stolen. As a precaution a blood sample was taken which revealed a blood alcohol concentration of $1.12\,g\,kg^{-1}$.

The car was found one day later in a backyard not far from the apartment. On the steering wheel only clear fingerprints of the owner were traced. The owner was confronted with this finding and confessed that he was the driver. He stuck to his claim concerning the alcohol intake after the accident and added that he did not drink before. As according to Widmark's formula the alleged volume of 200 ml US whiskey could only produce a blood alcohol concentration of 0.33–$0.68\,g\,kg^{-1}$ after 3.5 h of elimination, this could not be true. However, the postconsumption could not be excluded. Therefore, the respective concentration of maximally $1.03\,g\,kg^{-1}$ had to be subtracted. Recalculation resulted in an actual blood alcohol concentration of $0.49 - 0.89\,g\,kg^{-1}$ for the time of the accident. As the lower value is below the legal limit the driver would have been cleared of suspicion.

This was why a congener analysis of the stored blood sample was ordered by the court. **Table 4** demonstrates how the expert opinion was elaborated. Comparison of the actual congener concentrations of the blood sample with the expected values calculated with the correlation formulas of Table 3 shows that the real concentrations of methanol and propan-1-ol are higher than expected. As already known from the ethanol calculation (cf. above) the driver must have drunk also before the accident. This consumption can

Table 4 Comparison of analytical results and expected congener concentrations calculated from a postconsumption claim

Given times		
Preconsumption:	no statement	Accident: 00.30 h
Postconsumption:	01.00–03.00 h	Blood sample: 04.30 h
Body data		
Weight: 80 kg	Height: 180 cm	$r_{A\ med}$ to be taken
Drinking amounts claimed		
Preconsumption:	no alcoholic drinks	
Postconsumption:	US-Whiskey 'Jim Beam', 200 ml	
Congener content of 'Jim Beam' ($mg\,l^{-1}$)		
Methanol: 306	Isobutanol: 365	
Propan-1-ol: 50	3-Methylbutan-1-ol: 1320	
Blood-congener concentrations ($mg\,l^{-1}$)		
Methanol: 2.40	Isobutanol: 0.04	
Propan-1-ol: 0.25	3-Methylbutan-1-ol: 0.10	
Expected concentrations ($mg\,l^{-1}$) (cf. Table 3)		
Methanol: 1.05 ($s = 0.44$)	Isobutanol: 0.31 ($s = 0.09$)	
Propan-1-ol: 0.12 ($s = 0.07$)	3-Methylbutan-1-ol: 0.25 ($s = 0.04$)	

be responsible for the relatively high concentrations of these two congeners, because methanol is not metabolized as long as the blood ethanol concentration exceeds $0.4\,g\,kg^{-1}$ and propan-1-ol is eliminated much slower than the other fusel alcohols. On the other hand the actual concentrations of isobutanol and 3-methylbutan-1-ol are less than expected. The differences exceed 6 standard deviations altogether. This means that the claim can be excluded almost certainly. The court accepted this opinion and sentenced the accused.

Example 2 A quarter past midnight a man left a bar and started his car. When leaving the car park he scraped another car just entering the car park. He left his car and looked at the damage. When the other driver went to the bar in order to call the police, he left the scene with his car. The police searched for the damaged car. About two hours later the car was found in a suburban street. The driver was inside his car. He told the police that he had drunk only half a liter of beer during the last two hours before midnight in the bar mentioned. Because of the accident he was so frightened that he drove away randomly, stopped somewhere, opened the trunk and pulled five bottles of beer out of a crate and emptied them because of his excitement. The police examined the trunk. There was a crate containing 20 empty bottles of beer. Later on the police questioned the barkeeper, who stated that the driver had drunk 2.5 liters of beer during his stay in the bar. However, there remained doubts, because the other party of the accident was the son of the barkeeper.

The blood sample of the driver contained $1.21\,g\,kg^{-1}$ alcohol (and a second one taken 45 min later $1.13\,g\,kg^{-1}$). As a preconsumption of 2.5 liters of beer (instead of only half a liter as claimed by the driver) and an additional postconsumption of 2.5 liters would have produced a much higher blood ethanol concentration, either the statement of the barkeeper or the claim of the driver must be false. Calculating both alternatives with Widmark's formula did not allow differentiation. This was the reason for the court to order a congener analysis of the stored blood sample. The results are presented in **Table 5**. Comparison of the analytical results with the expected concentrations shows that the detected fusel alcohol concentrations are lower than those expected after the alleged postconsumption. The differences amount to 10 standard deviations altogether. On the contrary, there is a good correspondence with the expected fusel-alcohol concentrations calculated

Table 5 Comparison of analytical results and expected congener concentrations calculated from a postconsumption claim and a testimony, respectively.

Given times			
Preconsumption:	22.00–24.00 h	Accident:	00.15 h
Postconsumption:	00.30–02.00 h	Blood sample:	02.30 h
Body data			
Weight: 75 kg	Height: 176 cm	$r_{A\ med}$ to be taken	
Drinking amounts claimed			
Preconsumption:	Beck's beer, 500 ml		
Postconsumption:	Beck's beer, 2500 ml		
Drinking amount according to testimony			
Preconsumption:	Beck's beer, 2500 ml		
Congener content of 'Beck's Beer' ($mg\,l^{-1}$)			
Methanol: 9	Isobutanol: 9		
Propan-1-ol: 9	3-Methylbutan-1-ol: 62		
Blood-congener concentrations ($mg\,l^{-1}$)			
Methanol: 0.50	Isobutanol: 0.06		
Propan-1-ol: 0.21	3-Methylbutan-1-ol: 0.05		
Expected concentrations according to claim ($mg\,l^{-1}$)			
Methanol: 0.41 (s = 0.58)	Isobutanol: 0.19 (s = 0.11)		
Propan-1-ol: 0.35 (s = 0.05)	3-Methylbutan-1-ol: 0.35 (s = 0.05)		
Expected concentrations according to testimony ($mg\,l^{-1}$)			
Methanol: 0.57 (s = 0.28)	Isobutanol: 0.07 (s = 0.04)		
Propan-1-ol: 0.22 (s = 0.12)	3-Methylbutan-1-ol: 0.07 (s = 0.06)		

from the testimony. Therefore, the claim of the driver was disproved almost certainly, whereas the statement of the barkeeper could be confirmed. The driver was convicted by the court.

These examples show the information that is necessary for meaningful congener analysis: times of pre- and postconsumption, body data, amount and type of consumed beverages. If the brands of the beverages involved are known, they should be included in the analytical procedure. If only the types are known, their congener contents as listed in Table 2 can be used. There are, however, other limitations. The amounts of beverages consumed should exceed 1 l of beer, 0.5 l of wine and 60 ml of spirits. Spirits such as vodka or gin that are (nearly) congener-free are problematic, unless it is claimed that only such drinks were consumed and the blood sample contains congeners. The consumption time should be between 1 and 3 h and the time between end of drinking and blood sampling below 3 h. When these limitations became known to the public it was feared that the defense tactics would be altered, but so far they have not been.

As demonstrated most alcoholic beverages contain the same congener alcohols, but in different amounts. As the blood concentrations and their relations to each other are changed during elimination there is no obvious correspondence between the congener spectrum of a blood sample and the beverage consumed. This means that it is not possible to determine the type of beverage from a blood-congener analysis. Only if butan-2-ol is found can one conclude that a respective spirit must have been consumed. The congener spectrum of a urine sample corresponds much better to that of the beverage consumed. Such, additional urine samples can be useful on special occasions. However, there is no clear correlation between drinking amounts and urine concentrations.

The congener method is widely accepted by the German courts. Congener analyses are offered by nearly all German institutes of forensic medicine. In the Duesseldorf institute about 100 expert opinions based on this method are written each year. In around 75% of the cases the claim is almost certainly excluded, in around 14% the allegation is valued as improbable and in 10% as irrefutable. In the remainig cases the claim is confirmed. Astonishingly the opinions are also very often accepted by the accused. Often they make a confession during the court procedure. The author has been contacted more than once by an accused after conviction and assured that the conclusions exactly matched the real circumstances.

Although the congener method was presented more than once to the international public during the first years there was little interest. However, during the mid 1980s colleagues from northern and eastern Europe, and later on also from some western countries asked for cooperation. In 1987 an international workshop under the auspices of the International Committee on Alcohol, Drugs and Traffic Safety was held in Duesseldorf. It was found that the hit-and-run problem is not specific to Germany, which means that the method could also be advantageous in other legal systems.

Congener (especially methanol) findings can also be useful for the diagnosis of chronic alcohol abuse. As there are no analytical difficulties it has much to recommend it, especially for epidemiological studies. As an example the following observations may be mentioned. From forensic blood samples those with blood ethanol concentrations below $2.0\,g\,kg^{-1}$ and those with levels above $2.5\,g\,kg^{-1}$ were separated. The reason for this selection was the assumption that drivers with ethanol levels below $2.0\,g\,kg^{-1}$ may nearly represent the normal population. Levels above $2.5\,g\,kg^{-1}$ will only seldom be reached by normal persons and this was expected to contain a good numbers of alcoholics.

The findings confirmed this assumption. Concerning the methanol levels in the group of drivers with ethanol concentrations below $2.0\,g\,kg^{-1}$ there were striking similarities to the nondrinker group of the above mentioned experiments (Fig. 5). Even more interesting was the comparison of the chronic alcoholics with the drivers with ethanol concentrations above $2.5\,g\,kg^{-1}$ (Fig. 5). There was also a conspicuous correspondence indicating that at least 40% of these drivers must have alcohol problems, possibly even more than 70%.

See also: **Alcohol**: Blood; Interpretation.

Further Reading

Andréasson R (1986) *Erik Matteo Prochet Widmark*, pp. 1–132. Stockholm/Sweden: The Information Center for Traffic Safety.

Bonte W (1987) *Begleitstoffe alkoholischer Getränke. Biogenese, Vorkommen, Pharmakologie, Physiologie und Begutachtung*. Lübeck/Germany: Schmidt-Römhild.

Bonte W (1987) *Congener Alcohols and Their Medicolegal Significance*. Stockholm: Dalctraf.

Bonte W (1990) Contributions to congener research. *Journal of Traffic Medicine* **18**:5–15.

Bonte W (1991) Advances in congener analysis. In: Nagano T (ed.) *Proceedings of the First International Symposium on Advances in Legal Medicine*, pp. 344–348. Kanazawa, Japan: Tanaka Shobundo.

Bonte W, Kardel B and Eckrodt H (1988) A cryofocussing technique for head-space Gc with capillary columns. In: Jones GR and Singer PP (eds) *Proceedings of the 24th International Meeting of TIAFT*, pp. 340–343. Edmonton: Alberta Society of Clinical and Forensic Toxicologists.

Bonte W and Kühnholz B (1985) Pharmacokinetics and metabolism of fusel alcohols. In: Kaye S and Meier GW (eds) *Alcohol Drugs and Traffic Safety*, pp. 189–198. San Juan, Puerto Rico: US-Government Printing Office.

Bonte W, Sprung R, Rüdell E and Frauenrath C (1981) Forensic significance of congeners of alcoholic beverages. In: Goldberg L (ed.) *Alcohol Drugs and Traffic Safety*, pp. 1074–1081. Stockholm: Almqvist & Wiksell International.

Eichner A, Hoffmann G, Herrmann A, Rittner C and Wehner HD (1987) Calculation of elimination constants and distribution factors of congeners in individuals drinking brandy after beer consumption. In: Bonte W (ed.) *Congener Alcohols and Their Medicolegal Significance*, pp. 94–99. Stockholm: Dalctraf.

Freudenstein P and Bonte W (1987) Elimination kinetics of fusel alcohols in chronic alcoholics. In: Bonte W (ed.) *Congener Alcohols and Their Medicolegal Significance*, pp. 148–153. Stockholm: Dalctraf.

Gilg T, Scharschmidt A and vMeyer L (1987) Relations between congener and ethanol concentrations in blood after short-time intake of beer and wine. In: Bonte W (ed.) *Congener Alcohols and Their Medicolegal Significance*, pp. 116–125. Stockholm: Dalctraf.

Iffland R (1987) Congener-analyses of blood in legal proceedings. Possibilities and problems. In: Bonte W (ed.) *Congener Alcohols and Their Medicolegal Significance*, pp. 236–252. Stockholm: Dalctraf.

Jones AW (1988) Widmark's equation. Determining amounts of alcohol consumed from blood alcohol concentrations. *Drunken While Intoxicated Journal* 3:8–11.

Kühnholz B and Bonte W (1985) A method to determine fusel alcohols in blood samples. In: Kaye S and Meier GW (eds) *Alcohol Drugs and Traffic Safety*, pp. 781–788. San Juan: US-Government Printing Office.

vMeyer L, Steinbach T and Gilg T (1987) Contribution to the kinetics of congeners after drinking wine. In: Bonte W (ed.) *Congener Alcohols and Their Medicolegal Significance*, pp. 126–139. Stockholm: Dalctraf.

Schmutte P, Bilzer N and Penners B (1987) Some new aspects concerning the pharmacokinetics of congeners in human beings. In: Bonte W (ed.) *Congener Alcohols and Their Medicolegal Significance*, pp. 112–115. Stockholm: Dalctraf.

Urban R, Wolf M, Tutsch-Bauer E and Tröger HD (1987) Practical experiences with congener analysis. In: Bonte W (ed.) *Congener Alcohols and Their Medicolegal Significance*, pp. 253–267. Stockholm: Dalctraf.

Wehner HD, Hoffmann G and Azcarate L (1987) Distribution and elimination of congeners (n-propanol, iso-butanol) in a two-compartment model. In: Bonte W (ed.) *Congener Alcohols and Their Medicolegal Significance*, pp. 100–111. Stockholm: Dalctraf.

Interpretation

B Levine, Office of the Chief Medical Examiner, Baltimore, MD, USA
G W Kunsman, Bexar County Medical Examiner, San Antonio, TX, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0407

Introduction

Ethanol (alcohol) is, by far, the most commonly used and abused drug in modern society. In fact, alcohol is so commonly used it is seldom thought of as a drug at all, let alone a drug of abuse. Based on its frequency of use, it follows that the analysis of alcohol is the most frequently performed assay in forensic toxicology laboratories, both in the areas of postmortem toxicology and human performance toxicology. It also follows, then, that the results of these alcohol analyses are the most frequently employed laboratory results in criminal courts and the forensic toxicologist is frequently called on to provide an interpretation of these alcohol results.

The interpretation of alcohol results may range from determining its role in the cause of death in postmortem cases, differentiating antemortem alcohol consumption from postmortem alcohol production, to evaluating the effect of alcohol on driving performance in driving-while-intoxicated (DWI) cases. In general, however, the interpretation of alcohol results typically focuses more often on a discussion of its impairing effects on human performance, behavioral toxicity, than on its overt physiological toxicity. There are many factors that must be considered in the interpretation of alcohol results beyond its mere presence or the concentration at which it is found in a biological specimen.

The general effects of alcohol on human performance have been well established in both controlled scientific studies and epidemiological studies. These studies have allowed for the elucidation of a correlation between blood alcohol concentration (BAC) and performance effects. The degree and extent of impairment associated with a given blood alcohol concentration, however, may vary from that expected for a number of reasons. The level of impairment may be less than expected in individuals who have developed tolerance to the effects of alcohol. The level of impairment may be greater than expected in individuals who are naive to alcohol use, such as children and adolescents. The presence of other drugs may also alter the expected effects of alcohol on performance.

The forensic toxicologist may also be required to

interpret the antemortem or perimortem effects of alcohol based on a postmortem alcohol concentration. A significant issue that must be considered in the interpretation of alcohol results under these conditions is whether the measured concentration truly reflects the alcohol concentration at the time of death. This issue is significant because both postmortem increases and decreases in alcohol concentration have been reported.

This section will discuss the effect each of these factors plays on the interpretation of alcohol results.

Impairment

Ethanol is a short-chain aliphatic alcohol that is the natural product of sugar oxidation by yeast, a process referred to as fermentation. A small polar molecule, ethanol distributes throughout the total body water and, therefore, is found in all tissues and body fluids in proportion to their water content. As a pharmacologically active agent ethanol is classified as a central nervous system (CNS) depressant. Its mechanism as a CNS depressant may be through a direct action on the brain by dissolution of the molecule in the neuronal plasma membrane. This leads to a disruption of cellular functions by disordering the lipids of the cellular membrane. In addition, there is significant evidence for the existence of an ethanol receptor. This receptor is thought to be part of the γ-aminobutyric acid (GABA) receptor complex, therefore, ethanol binding to the receptor enhances chloride ion movement into the cell. The mechanism for the reinforcing and mood-elevating effects of ethanol is unclear.

The behavioral effects of ethanol are consistent with a generalized central nervous system depression. The degree of depression and, hence, the degree or extent of impairment exhibits a dose-dependent relationship. The venous blood alcohol concentration and the breath alcohol concentration are also highly correlated with the type and extent of impairment associated with alcohol use. In general, venous blood or breath is sampled for analysis when the use of alcohol is suspected as the source of observed or measured impairment. The psychomotor impairment associated with alcohol use has been well documented in both laboratory studies and in actual driving studies, both closed-course and on-the-road driving. Although alcohol has been shown to impair a number of different tasks that rely on cognition and motor skills, the task of greatest interest and which most studies are directed towards measuring is the routine, yet complex psychomotor task of driving an automobile. Even laboratory studies that are designed to measure the most basic and simple behavioral tasks are utilized in an effort to understand how each element of the driving task, e.g. reaction time, tracking ability, divided attention, etc., is impaired by the use of alcohol.

The United States National Safety Council's Committee on Alcohol and Drugs released a resolution in 1971 regarding alcohol impairment in which they stated that no individual, no matter what his previous experience with alcohol, is not unimpaired in his driving performance if his BAC is $0.08\,g\,dl^{-1}$ or greater. Although the presence of impairment at the $0.08\,g\,dl^{-1}$ blood alcohol level is generally accepted, a number of studies have demonstrated that the degree of impairment at this alcohol level can be highly variable between individuals. This variability is based on a number of factors including, but not limited to, the subject's degree of experience with alcohol, whether their blood alcohol concentration is rising or falling, and their level of training and experience with the task on which they are tested. There are also a number of studies that suggest some behaviors may be impaired at blood alcohol concentrations as low as $0.02\,g\,dl^{-1}$. It is important to note that the type and degree of impairment is more variable between subjects at lower alcohol concentrations than it is at higher concentrations. Also, a greater proportion of people in any test population are impaired at higher blood alcohol concentrations, especially greater than $0.05\,g\,dl^{-1}$, than at blood alcohol concentrations below that level.

The general effects of alcohol on selected human performance measures are briefly summarized as follows. This summary is not a comprehensive review of every type of behavior affected by alcohol, but provides an overview of the types and degree of behavioral effects associated with alcohol consumption.

- A reduction in visual acuity and impaired peripheral vision has been documented at a BAC of $0.07\,g\,dl^{-1}$ and the extent of impairment increased with increasing BAC.
- Low doses of alcohol have been noted to decrease the subject's sensitivity to taste and smell.
- A decrease in pain sensitivity is experienced at a BAC of $0.08\,g\,dl^{-1}$ and increases with increasing concentration.
- An altered time sense, manifested as the slowing down of the passage of time, has been documented following low and moderate alcohol doses (BAC less than $0.08\,g\,dl^{-1}$).
- Choice Reaction Time (CRT) is impaired at concentrations of $0.05\,g\,dl^{-1}$. CRT is a motor performance test used to evaluate the integrity and function of motor pathways. Reaction time tasks typically use a button press in response to a critical stimulus to measure a subject's latency to respond to that stimulus. CRT measures sensorimotor

performance by selecting a single stimulus from among a number of alternatives. The task has both a recognition time component and a motor response component. The impairment measured consists of both an increased latency to respond to the stimulus and a decrease in the accuracy of the response.

- Tests of hand–eye coordination have measured deficits in performance at concentrations of $0.05\,\mathrm{g\,dl^{-1}}$ and greater.
- Performance on vigilance tasks has been shown to be impaired at a BAC of $0.06\,\mathrm{g\,dl^{-1}}$. Vigilance tasks are tests of intellectual function and measure the ability to discriminate a specific signal from among a group of choices. Such tasks provide a measure of an individual's ability to recognize specific information. Impaired vigilance is probably a reflection of the drowsiness associated with alcohol consumption.
- An increase in body sway, as measured by the Romberg test and a laboratory-based performance device referred to as a wobble board, has been documented at a BAC of $0.05\,\mathrm{g\,dl^{-1}}$. The body sway degrades to staggering and reeling as the blood alcohol concentration increases.
- The standardized field sobriety test (FST) employed by many law enforcement agencies is a battery of psychomotor performance tests that is utilized at the roadside to document the impairment of drivers suspected of DUI (driving under the influence)/DWI. The three tests that constitute the standardized field sobriety test are the one leg stand (OLS), the walk and turn (WAT), and horizontal gaze nystagmus (HGN).

  - The OLS is a divided attention task with two distinct stages, instruction followed by balancing and counting. The Officer notes impairment by the number of errors, referred to as clues, made on the test. There are a maximum number of five clues on this test. If the individual tested scores two or more errors or is unable to complete the test there is a 65% predictability that the individual's BAC is $0.10\,\mathrm{g\,dl^{-1}}$ or greater.
  - The WAT is a divided attention task with two distinct stages, instruction and walking. There are a maximum number of nine clues on this test. If two or more clues are scored or the subject is unable to complete the test there is a 68% predictability that the individual's BAC is $0.10\,\mathrm{g\,dl^{-1}}$ or greater.
  - The HGN test measures central nervous system motor pathway function. Nystagmus is a normal phenomenon that is not caused by alcohol, but is enhanced by alcohol. HGN is the most sensitive test in the FST battery to the impairing effects of alcohol. The officer observes the suspect's eyes for the presence of smooth tracking and the onset of nystagmus. Studies have shown that the earlier that nystagmus occurs (the shorter the angle from directly in front of the subject) the greater the blood alcohol concentration. In fact, there appears to be a dose–response relationship between BAC and the angle of onset of nystagmus. There are a maximum number of six clues on the HGN, three per eye. If four or more clues are scored there is a 77% predictability that the individual's BAC is $0.10\,\mathrm{g\,dl^{-1}}$ or greater.

- Standard intelligence tests, such as the Wechsler Adult Intelligence test, have measured impaired cognition associated with alcohol use. This impairment increases with increasing dose.
- Most tests of driving skill, both on the road tests and driving simulators, show impairment at a BAC of $0.05\,\mathrm{g\,dl^{-1}}$. Epidemiological studies suggest an association between impaired driving performance and alcohol use. The majority of these studies show that 40%–60% of all fatally injured drivers have blood alcohol concentrations of $0.10\,\mathrm{g\,dl^{-1}}$ or greater and 30%–40% of those have a BAC in excess of $0.15\,\mathrm{g\,dl^{-1}}$. It is important to note that a cause-and-effect relationship cannot be established by retrospective studies, therefore, it is not possible to state, based on these studies, that the presence of alcohol was the causitive factor in the accidents.
- The broad range of alcohol effects on behavior may be loosely categorized by BAC as follows:

  - $<0.05\,\mathrm{g\,dl^{-1}}$ (low dose) – increased talkativeness, mild excitement, decreased attention, decreased inhibitions, and some minor motor skills impairment in some individuals;
  - $0.05-0.10\,\mathrm{g\,dl^{-1}}$ (moderate dose) – talkative, cheerful, loud, boisterous then sleepy, increased confidence, increased risk-taking, and impaired psychomotor skills (tracking, vigilance, divided attention, reaction time, etc.);
  - $0.10-0.30\,\mathrm{g\,dl^{-1}}$ (elevated dose) – nausea and vomiting may occur followed by lethargy, ataxia, slurred speech, diplopia, staggering gait, disorientation, and grossly impaired psychomotor skills;
  - $0.30-0.50\,\mathrm{g\,dl^{-1}}$ (high dose) – stupor, visual impairment, marked decreased response to stimuli (even painful stimuli), and marked muscular incoordination. Coma and eventually death due to respiratory depression are generally accepted to occur at a BAC greater than $0.40\,\mathrm{g\,dl^{-1}}$ in nonalcohol-dependent individuals. Note that highly tolerant individuals

(e.g. chronic alcoholics) may not experience or appear to experience many of the more serious effects associated with high blood alcohol concentrations, but all individuals experience the cognitive and judgment impairing effects of alcohol at blood concentrations greater than $0.08\,g\,dl^{-1}$.
  - The general effects of alcohol on behavior in relation to a range of blood alcohol concentrations have been delineated in table format in a number of reference books to facilitate an understanding of how BAC correlates to performance impairment.

- A large number of subjective tests have been used in association with the behavioral tasks described above. In these tasks the subject self-reports their mood, feelings, and impressions using a quantifiable scale, such as Self-rated Mood Scales, the Hopkins Symptom Checklist, the Cornell Medical Index, and other self-rated performance evaluations. These subjective tests indicate that with increasing BAC subjects typically feel elated, friendly and vigorous. As their BAC decreases after reaching its peak concentration they generally feel anger, depression and fatigue. Another common finding on such subjective tests is that subjects commonly underrate the extent of their psychomotor impairment.

Many of the tasks utilized in these studies show that behavioral tolerance may develop with repeated ethanol use. Tasks learned under the influence of alcohol are often performed better when repeated at that blood concentration than when performed in the absence of alcohol. Generally, the more complex the task the more significant the impairment measured at lower doses of alcohol. There is also a large between-study and between-subject variability in the type and extent of impairment noted. This is especially true when the blood alcohol concentration in subjects is below $0.08\,g\,dl^{-1}$. When interpreting the results of behavioral studies it is important to recognize that the reported results often refer only to some of the subjects tested and summary results typically are indicative of population tendencies and do not reflect absolute measures of behavioral effects.

In summary, even at low levels, ethanol disrupts performance and can interfere with complex activities such as driving. It generally causes feelings of happiness and reduces the ability of aversive events, such as pain, to control behavior. As the blood alcohol concentration increases the degree of impairment also increases and may eventually result in a loss of consciousness and finally death. The effects of alcohol on behavior are generally more pronounced and pleasurable while the blood alcohol levels are rising than while they are falling. There are also a number of behaviors that exhibit tolerance to the impairing effects of ethanol. This tolerance may be a consequence of both increased metabolism and learning.

A Brief History of Alcohol and the Law

The ability of alcohol to impair psychomotor performance and to produce behavioral changes has been well documented throughout history. The use and especially abuse of alcohol has always had a negative impact on society, however, the advent of the industrial age and the invention of the automobile have rendered these effects of even greater significance. As early as 1843 the New York Central Railroad prohibited employees to drink while on duty. In 1910 the New York City traffic code noted that the misuse of alcohol was a factor in traffic safety.

The increasing mechanization of American industry and the increasing use of automobiles was also accompanied by an ever-increasing awareness of safety issues, not only in factories, but also on the roads and in the home. The formation of the National Council for Industrial Safety in 1912, which became the National Safety Council in 1914 was a significant step in the promotion of the safety movement in the United States. By 1924 the National Safety Council expanded its interests to include highway safety, and therefore, by implication, to the effects of alcohol on driving. The work of this organization has been continued and expanded by the National Highway Traffic Safety Administration (NHTSA).

The scientific support for this safety movement did not begin until the early 1920s when Professor Widmark, from the University of Lund in Sweden, developed a protocol for physicians to follow in the evaluation of drivers suspected of driving under the influence (DUI) of alcohol. From that point forward, the role of the scientist and scientific evidence gained a more and more important role in the relationship between alcohol and the law. The first law passed in the United States directed at drinking and driving was the Connecticut Motor Vehicle Law passed in 1924; that law stated that no one who has been drinking ought to be allowed to operate a motor vehicle. In 1935, Richard Holcomb and the Northwestern University Traffic Institute initiated a three-year study called the Evanston Study. This study reported on 270 drivers hospitalized after involvement in automobile accidents in Evanston, IL. The Evanston police tested 1750 drivers for their blood alcohol concentration and their breath alcohol concentration using Rolla Harger's recently invented Drunkometer

over the same three-year period. In 1938, Holcomb reported that the chances of having an accident increased dramatically with the presence of any alcohol in the blood to the extent that each $0.02\,g\,dl^{-1}$ rise in blood alcohol resulted in a doubling of the risk of accident.

The results of this study and the subsequent recommendations included in the joint 1938 statement issued by the Committee to Study Problems of Motor Vehicle Accidents (a special committee of the American Medical Association) and the Committee on Alcohol and Other Drugs formed the basis for the first legislation in the United States making DUI an offense. This legislation was passed in Indiana in March, 1939 and in Maine in April, 1939. The recommendations of these two committees also formed the basis for the Chemical Tests Section of the Uniform Vehicle Code published by the National Committee on Uniform Traffic Laws and Ordinances in 1946. In 1953, Implied Consent legislation was passed in New York State and was soon included in the Uniform Vehicle Code. Implied consent laws have subsequently been passed in all fifty States. The implied consent legislation provides that, as a condition precedent to being issued a driver's license, an applicant agrees, by implication, to submit to a chemical test in any case in which he is suspected of DUI. Refusal to submit to the test results in the temporary loss of driving privileges.

In 1958, a Symposium on Alcohol and Road Traffic held at Indiana University issued a statement that a BAC of $0.05\,g\,dl^{-1}$ will definitely impair the driving ability of some individuals. As the BAC increases, a higher proportion of individuals will become impaired until a $0.10\,g\,dl^{-1}$ is reached, at which point all individuals are definitely impaired. The Committee on Alcohol and Drugs in 1960 released a statement recommending that DUI laws be amended to reflect a $0.10\,g\,dl^{-1}$ BAC as presumptive evidence of guilt; prior to this date, the presumptive concentration defined in most State laws was a $0.15\,g\,dl^{-1}$. The Uniform Vehicle Code was amended to reflect this recommendation in 1962.

The Grand Rapids Study, published in 1964 by Indiana University researchers, confirmed the results of the Evanston study and also stated that drivers with blood alcohol concentrations greater than $0.04\,g\,dl^{-1}$ tend to have more single vehicle accidents that are also more severe than do sober drivers. Another significant finding from this study was that accident-related factors other than alcohol decreased in significance when the driver's BAC was greater than $0.08\,g\,dl^{-1}$ and that accident involvement increased rapidly when the driver's BAC was greater than $0.05\,g\,dl^{-1}$. Drivers with BAC levels in the range 0.04–$0.08\,g\,dl^{-1}$ had a greater risk of accident involvement, but alcohol was not necessarily more significant than other risk factors.

The passage of the National Highway Safety Act in 1966 began the era of Federal intervention in the drinking and driving problem in earnest. The NHTSA, part of the newly created Department of Transportation, submitted a report to Congress in 1968 detailing how the problem of the drunken driver was being addressed. In 1971, NHTSA released a statement that no individual, no matter what his previous experience with alcohol, is not unimpaired in his driving performance if his BAC is $0.08\,g\,dl^{-1}$ or greater. The last three decades have seen a continued proliferation of regulations and legislation concerning the drinking and driving problem. Although the legal limit for driving while impaired has remained at a $0.10\,g\,dl^{-1}$ BAC in most states of the United States, a number of jurisdictions have begun to lower this limit to a $0.08\,g\,dl^{-1}$ BAC with some setting even lower limits for individuals younger than the legal drinking age. Law enforcement agencies have contributed to these efforts through the development and implementation of programs directed at reducing the number of impaired drivers through increased intervention and education.

The standardized field sobriety test and the drug evaluation and classification (DEC) program are two of the intervention programs that have been developed. The DEC program is primarily directed toward the training of police officers as drug recognition experts (DRE) in an effort to provide a mechanism for obtaining compelling evidence that a driver was impaired, specifically by a drug other than or in addition to alcohol, at the time of the stop. The FST, the individual components of which are described above (OLS, WAT, and HGN), is a series of pyschomotor tests used to measure impairment at the roadside following a traffic stop. The standardized field sobriety tests were developed in the 1970s through funding provided by NHTSA and have been standardized through laboratory studies and validated in field studies. Although drugs other than alcohol may impair the behaviors evaluated with the FST, the tasks have primarily been validated against a blood alcohol concentration of $0.1\,g\,dl^{-1}$.

Alcohol Effects on Children and Adolescents

A large volume of data has been accumulated on the effects of ethanol on adults. Much of this information has been gathered from controlled scientific studies in which adult subjects were given a dose or multiple

doses of ethanol and then asked to perform a task or series of tasks. Ethical considerations prevent the conducting of these types of experiments with children or adolescents. Although there are age restrictions in most countries for the legal consumption of ethanol, there have been a number of reports where children or adolescents have accidentally or intentionally consumed alcoholic beverages. These reports have indicated several things. For example, children appear to be more sensitive to the effects of ethanol than adults, that is, lower blood ethanol concentrations produce more significant toxicity in children than in adults. One study documenting this fact revealed that young teenagers were in a coma with positive pain reaction at an average blood ethanol concentration of $0.15\,g\,dl^{-1}$ and in a coma with no pain reaction at a blood ethanol concentration of $0.19\,g\,dl^{-1}$. These concentrations in adults would not be expected to produce a loss of consciousness.

Alcohol Tolerance

One of the most significant factors complicating the interpretation of blood ethanol concentration is the phenomenon of tolerance. Tolerance is a condition in which a decreased response to the effects of alcohol, or other drug, is acquired in the face of repeated exposure to alcohol or that other drug. The consequence of the development of tolerance is that it becomes necessary to successively continue to increase the dose of alcohol to achieve an equal pharmacological effect or duration of action. Tolerance may also be thought of as that condition in which a given dose of alcohol fails to produce the same effect or duration of action as a previous equivalent dose of alcohol. There are several types of tolerance as that phenomenon applies to ethanol.

Mellanby first described an acute tolerance to ethanol in 1919. Using dogs as his model, he showed that at a given blood ethanol concentration, intoxication was less severe during the descending portion of the blood ethanol concentration versus time curve than during the ascending portion. This acute tolerance to ethanol has become known as the 'Mellanby effect' and has subsequently been verified by a number of researchers. The Mellanby effect has been observed with tasks that measure psychomotor performance, cognitive performance, and on the subjective effects associated with alcohol use. The demonstration of acute tolerance depends on the range of blood ethanol concentrations studied and the tests employed to assess tolerance. Studies conducted to evaluate the development of acute tolerance have utilized the following experimental designs:

1. the measurement of the change in performance at the same blood ethanol concentration on the ascending and descending limbs of the blood ethanol concentration versus time curve;
2. the determination of the blood ethanol concentration at the onset of measurable impairment and when that impairment is no longer measurable;
3. the measurement of task performance when the blood ethanol concentration is maintained constant.

In addition to the development of acute tolerance to ethanol, there is also an acquired or chronic tolerance to ethanol. Acquired tolerance has been demonstrated by (1) comparing performance on specific tasks between light and heavy users of ethanol and (2) the development of experimentally acquired tolerance under laboratory conditions. A number of studies have demonstrated that heavy drinkers exhibit less alcohol-induced psychomotor impairment on the same tasks than light drinkers. It is less certain whether acquired tolerance develops to the impairing effects of ethanol on memory and cognitive function. A significant limitation to the development of a more comprehensive understanding of acquired tolerance is that controlled scientific studies of the effects of ethanol on behavior in human subjects are generally limited to blood ethanol concentrations no greater than $0.10\,g\,dl^{-1}$. Anecdotal reports, however, indicate that chronic alcoholics can have much higher blood ethanol concentrations without displaying overt symptoms of intoxication than can light drinkers.

Metabolic tolerance to ethanol also develops. Chronic alcoholics typically eliminate ethanol from the blood at higher rates than the occasional or social drinker does. The microsomal ethanol oxidizing system, which may become involved in ethanol metabolism at high blood alcohol concentrations and in chronic alcohol use, is inducible and may account for the increase in alcohol metabolism.

Drug Interactions with Ethanol

Another significant factor complicating the interpretation of ethanol concentrations is the co-administration of other drugs or chemicals. Alcohol can affect or be affected by other drugs both in terms of pharmacokinetics and pharmacodynamics. These interactions can be additive, synergistic, potentiating or antagonistic. An additive effect indicates that the total effect of a drug combination is the sum of the effects of the individual drugs. A synergistic effect means that the total effect of the drug combination is

greater than the sum of the effects of the individual drugs. Potentiation is defined as an increase in the effect of a toxic substance acting simultaneously with a nontoxic substance. Antagonism refers to the canceling of effects of one drug by the simultaneous administration of another drug. The following is a summary of some of the major drug interactions with ethanol that have been characterized.

Amphetamines

Amphetamines and other sympathomimetic amines may antagonize the depressant effects of ethanol, mainly by offsetting the fatigue produced by ethanol. Stimulants, in general, appear to diminish or negate the behavioral effects of ethanol on well-learned tasks, but have no impact on the impairing effects of alcohol on newly learned or unfamiliar behavioral tasks. The interactive toxic and behavioral effects of alcohol and stimulants are difficult to evaluate and may be dependent on a number of factors. Some of these factors are the relative dose of each drug, the relative time-frame of drug use, the complexity of the behavior being evaluated, the subject's experience with that behavioral task, and the subject's experience with or degree of tolerance to the drugs.

Antianxiolytics

A number of studies have been performed documenting the combined effects of ethanol and benzodiazepines. Among the most significant interactions is the resultant increased blood diazepam concentration following the co-administration of ethanol. Although such an increase has not been evaluated with most benzodiazepines, it is very likely to occur based on the structural similarities among the various members of the benzodiazepine class of drugs. The N-dealkylation and hydroxylation phase I reactions of benzodiazepine metabolism are inhibited by acute doses of ethanol. Benzodiazepines, however, do not affect alcohol dehydrogenase activity. Studies evaluating the behavioral effects of the combination of benzodiazepines and alcohol indicate an additive depressant effect on most measures of performance. In general, the behavioral effects of the benzodiazepines are very similar to those of ethanol and the two drugs in combination exacerbate the overt effects and apparent intoxication of each drug alone. Conversely, buspirone, a nonbenzodiazepine anxiolytic agent, does not potentiate the effects of low or moderate doses of ethanol.

Barbiturates

A synergistic effect in CNS depression is seen when alcohol and barbiturates are co-administered. In general, the behavioral effects of the barbiturates are very similar to those of ethanol and the two drugs in combination exacerbate the overt effects and apparent intoxication of each drug alone. Acute ethanol intoxication inhibits barbiturate metabolism, thereby increasing barbiturate concentrations and increasing their associated toxicity.

Cocaine

The combined use of ethanol and cocaine results in the formation of a unique metabolite, cocaethylene, by means of a transesterification process that occurs in the liver. The half-life of cocaethylene is slightly longer than cocaine, it is more toxic than cocaine, but it exhibits the same type and degree of CNS stimulation as cocaine. Therefore, the overall toxicity due to cocaine is increased when it is used in combination with ethanol. Only a limited number of studies on the combined behavioral effects of cocaine, a CNS stimulant, and ethanol, a CNS depressant, have been conducted. The performance-enhancing effect of cocaine noted in most studies appears to be the result of the stimulant's ability to reverse the effects of fatigue. Cocaine and ethanol interaction studies have shown that the addition of cocaine to ethanol does not enhance the impairing effects of ethanol on performance, but either attenuates the impairment resulting from ethanol consumption or leaves ethanol-induced impairment unchanged. The performance-enhancing effect of cocaine in these studies has been measured to last for a number of hours, appears to occur only in well-learned behaviors, and is most significant in fatigued subjects.

Histamine-2 antagonists

Histamine-2 antagonists such as cimetidine and ranitidine are commonly prescribed drugs used to treat peptic ulcer or excess stomach acid production. These drugs inhibit gastric alcohol dehydrogenase and, as a result, increase the bioavailability of ingested ethanol. These drugs also inhibit the cytochrome P_{450} microsomal enzyme system, which could affect ethanol metabolism by the microsomal enzyme oxidizing system.

Marijuana

At high doses, cannabis acts like a hallucinogen, but at the low doses commonly used in North America, the drug is reported to cause a pleasurable high that may take several trials to experience and can usually be turned off at will. Most performance deficits associated with marijuana use appear to be due to a

lack of motivation and an inability to attend to a task. The impairing effects are generally slight and measurable only in some of the individuals tested. The interaction of ethanol and marijuana is presumably additive, but is difficult to evaluate due to the high degree of inter-subject variability in the behavioral effects associated with marijuana use.

Opiates

The opiates are CNS depressants that produce analgesia, euphoria, sedation, respiratory depression, miosis, nausea and emesis. When alcohol and members of the opiate class are co-administered the CNS depression and behavioral impairment are, at minimum, additive. Acute doses of ethanol also lead to the decreased hepatic metabolism of opiates such as methadone and propoxyphene. This inhibition results in an increase in the blood concentration of the parent drug and an associated increase in the behavioral and toxic effects of that drug. The administration of opiates, however, has no apparent effect on the metabolism of ethanol.

Tricyclic antidepressants

Tricyclic antidepressants increase catecholamine neurotransmitter concentrations in the synaptic junction by blocking their neuronal reuptake. Tricyclic antidepressants are extensively metabolized in the liver and the acute ingestion of ethanol inhibits this metabolism resulting in increased blood concentrations and a greater risk of toxicity. The tricyclic antidepressants exert a profound sedative effect that is additive to the sedating effects of ethanol.

Decreases in Ethanol Concentration

The interpretation of alcohol concentration is also complicated by the potential loss of alcohol from biological specimens in storage. There are three mechanisms by which this decrease in concentration may occur over time: (1) evaporation; (2) oxidation; and (3) microbial action.

Evaporation

Since ethanol is a volatile substance, evaporation loss from biological specimens may occur over time if the specimen is not collected and stored properly. Obviously, if the specimen container is improperly sealed ethanol will be lost from the sample. This is especially true if the specimen is stored at room temperature or a slightly elevated temperature. When a specimen is collected for subsequent alcohol analysis it is important that there be minimal air space (headspace) between the top of the specimen and the lid. If too large an air space exists, the vapor pressure of ethanol will allow the movement of ethanol from the specimen into the headspace, with the eventual release of the vapor when the container is opened.

Oxidation

The *in vitro* oxidation of ethanol to acetaldehyde has been reported. This process is oxyhemoglobin-mediated and uses oxygen in the blood and from the air that is in contact with the stored blood specimen. In general, the amount of ethanol loss is minimal and is limited to approximately $0.04\,g\,dl^{-1}$.

Microbial action

Certain microorganisms can use ethanol as a substrate for metabolism. This is an aerobic process that is facilitated by the volume of air in the headspace above the sample. Strains of the bacteria *Serratia marcescens* and *Pseudomonas* sp. have been isolated from blood specimens in which ethanol loss has been documented.

Increases in Ethanol Concentration

The interpretation of alcohol concentration is also complicated by the potential increase in alcohol concentration in stored biological specimens. An increase in alcohol concentration in specimens collected from living subjects typically occurs only during collection, but can occur prior to, during and subsequent to collection in postmortem specimens. There are two mechanisms by which an increase in concentration may occur: (1) contamination; and (2) postmortem ethanol formation.

Contamination

Ethanol concentrations can be spuriously increased by external contamination of the specimen. Cleansing the collection site with an antiseptic containing ethanol is an obvious source of contamination. Alcohol-free antiseptics are available and should be used when collecting specimens for alcohol analysis. This is essentially the only way in which an increase in the ethanol concentration occurs in the specimens collected from living individuals. Although it is possible for bacterial colonies in the blood to produce ethanol as a byproduct of glucose metabolism healthy individuals do not have a sufficient number of bacterial colonies to produce measurable ethanol concentrations.

A number of additional sources for the possible external contamination by ethanol exist for postmortem specimens. In trauma cases for example, blood from the heart, a common site of postmortem

blood collection, may be contaminated by stomach contents. If there is any residual ethanol remaining in the stomach contents, this will cause an artificial increase in the heart blood ethanol concentration when the stomach contents come in contact with the heart. In these cases, blood from a peripheral site, away from the site of trauma, should be collected and analyzed for ethanol.

Embalming fluid may also be a source of ethanol contamination if the specimens are not collected prior to embalming. If the issue of ethanol consumption arises after a body is embalmed, it is recommended that some of the embalming fluid be obtained and analyzed for the presence of ethanol to determine its contribution to the postmortem blood concentration, although most embalming fluids do not contain ethanol. Autopsy specimens may also be contaminated during collection if syringes are flushed with alcohol between sample collections and if the same syringe is used for the collection of specimens from different decedents. A large number of literature reports have documented these sources of ethanol contamination of biological specimens. Due to these reports, the incidence of external contamination has been significantly decreased owing to an increased awareness of the potential problems of using ethanol swabs, reusing syringes for sample collection, and the importance of collecting autopsy specimens prior to embalming.

Postmortem ethanol formation

A critical component in the interpretation of postmortem blood ethanol concentrations is the issue of whether the measured ethanol resulted from the consumption of alcoholic beverages before death or from the production of ethanol after death. A variety of aerobic and anaerobic bacteria, yeast and molds can, under proper conditions, produce ethanol. A number of substrates can be converted into ethanol by these microorganisms. Glucose is the primary substrate that may be converted into ethanol, therefore, any tissue with high concentrations of glucose or glycogen is susceptible to postmortem alcohol production. Blood, liver and muscle are examples of specimens with high sugar concentrations in which significant concentrations of ethanol attributed to postmortem formation have been measured. Conversely, urine and vitreous humor are ordinarily free of the combination of glucose and microorganisms necessary to produce ethanol and are, therefore, excellent samples for evaluating the role of postmortem ethanol formation. Other substrates for ethanol production include lactate, ribose and amino acids. The mechanism of ethanol production from sugar is glycolysis, which is the first step in the normal breakdown of glucose.

Postmortem ethanol formation can occur in the body between death and specimen collection or can occur after the specimens are collected, but prior to analysis. The first condition is difficult to control since body recovery may be delayed for days or even weeks. The prevention of ethanol production after the specimens are collected can be accomplished by performing several simple procedures. Following collection, the specimens should be sent to the laboratory as soon as possible. At the time of collection blood specimens should be treated with a preservative and an anticoagulant. Sodium fluoride is the most common preservative and potassium oxalate is the most common anticoagulant used; both compounds are found in the gray top Vacutainer® tube. The addition of 1–2% (w/v) is generally sufficient to inhibit most microbial activity. Specimens should also be refrigerated upon receipt by the laboratory and kept refrigerated until analysis. For long-term storage of the specimens following analysis, frozen storage is the preferred mechanism for preserving specimens and preventing ethanol formation.

There are a number of factors that can or should be considered when determining whether measured ethanol occurred due to antemortem consumption of alcohol or microbial activity: (1) the decedent's drinking history; (2) notation of the signs of putrefaction; (3) the results of the analysis of multiple specimens; and (4) the production of volatile compounds in addition to ethanol.

Case history Witnessed drinking by the decedent prior to death is obviously significant. Although there is often an underestimation of the amount of ethanol consumed, the observation that the individual was drinking is usually reliable. Unfortunately, drinking history immediately prior to death is often unavailable, especially in the case of unattended deaths.

Signs of putrefaction Although conditions for a body to putrefy or decompose may vary tremendously, there are a number of common characteristics of a decomposed body. The most striking trait is the foul odor associated with the body. Bloating, discoloration, and skin slippage are also common features associated with decomposition. Insect infestation, such as maggots, is frequently present in decomposed bodies and is often helpful in ascertaining the length of time that an individual has been dead. When signs of putrefaction are present, postmortem production of ethanol must be considered as possible if not probable. Unfortunately, the amount of ethanol produced is highly variable between decedents; two bodies kept in the same conditions for the same length of time can produce widely different amounts of

ethanol. This issue is further complicated if the individual had actually been drinking prior to death. In that scenario, the postmortem alcohol measured might be due to both antemortem consumption and postmortem formation.

Results of multiple specimen analysis When ethanol is absorbed by the body, it distributes according to the water content of each tissue or fluid. The greater the water content, the greater is the ethanol concentration in that tissue or fluid. Since the water content of a fluid or tissue is relatively constant, once the ethanol has reached equilibrium, there is a predictable relationship between the ethanol concentration in the various tissues and fluids. Much of the work establishing these relationships has compared the ethanol concentration in blood to that of other fluids and tissues. For example, the vitreous humor and cerebrospinal fluid ethanol concentrations will typically be higher than the blood ethanol concentration after equilibrium has been reached. On the other hand, liver and brain will typically have lower ethanol concentrations than the blood ethanol concentration after equilibrium.

One advantage to postmortem toxicologic analysis is the availability of a wide variety of fluids and tissues that can be collected at autopsy. Since the analysis of ethanol in tissues is straightforward, multiple analyses for ethanol can readily be performed. The distribution of ethanol between these specimens can provide a strong indication as to whether the measured ethanol resulted from drinking or decomposition. For example, one approach to multiple specimen analysis is to analyze blood, vitreous humor and urine. In the postabsorptive phase of alcohol distribution, the vitreous humor to blood ethanol concentration ratio is about 1.2 and the urine to blood ethanol concentration ratio is about 1.3, although there are wide variations in these ratios. If the measured postmortem ethanol concentrations yield similar ratios to those established for these specimens, then it is reasonable to conclude that the measured ethanol resulted from drinking. Vitreous humor and urine are two specimens relatively resistant to the putrefactive process and thus, are not sites generally associated with postmortem ethanol formation. If the blood ethanol concentration is positive and the vitreous humor and urine ethanol concentrations are negative, this is a strong indication that the ethanol concentration in the blood is the result of decomposition. One study of postmortem alcohol production showed that a blood ethanol concentration of $0.01\,\mathrm{g\,dl^{-1}}$ was associated with a negative urine or vitreous humor ethanol concentration 46% of the time. When the blood ethanol concentration was $0.04\,\mathrm{g\,dl^{-1}}$, this percentage decreased to 8%. These results support the analysis of multiple postmortem specimens to evaluate the role of postmortem alcohol production and the fact that urine and vitreous humor are excellent samples for this purpose.

A number of studies have been performed that describe the production of ethanol in postmortem blood. These studies can be summarized by the following conclusions.

1. When ethanol is produced postmortem, the ethanol concentration is usually less than $0.07\,\mathrm{g\,dl^{-1}}$.
2. Production of ethanol concentrations greater than $0.10\,\mathrm{g\,dl^{-1}}$ has been reported.
3. The production of ethanol due to decomposition is variable and is dependent on the species of microorganism present, the available substrate, and the temperature and other environmental conditions.

Although urine has been shown to be generally immune to the effects of *in vitro* production of ethanol, several studies have indicated that the combination of glucose in the urine, a condition that often occurs in diabetics, and a *Candida albicans* infection can result in the production of large amounts of ethanol. Both components are required for ethanol production to occur and it will not occur in the absence of either the glucose or the microorganism. This can be demonstrated in the laboratory by performing serial analyses of the urine for ethanol over several days and observing the increase in ethanol concentration over time.

Production of other volatiles Microorganisms that produce ethanol may also be capable of producing other volatile substances, one of which is acetaldehyde. However, since acetaldehyde is also a metabolite of ethanol, its identification in biological specimens cannot be used as a marker for decomposition ethanol production. One other volatile commonly seen as a putrefactive product is n-propanol. n-Propanol is not identified in individuals drinking alcoholic beverages and, therefore, is a good marker for decomposition ethanol formation. One caution in the use of this alcohol as a marker of postmortem ethanol formation is that many laboratories use n-propanol as an internal standard for the analysis of ethanol by gas chromatography. Although the concentration of n-propanol added as an internal standard far exceeds the amount of n-propanol produced (generally $<0.01\,\mathrm{g\,dl^{-1}}$ produced in blood) and, therefore, would not significantly affect the ethanol quantitation, the use of n-propanol as an internal standard would mask the presence of n-propanol in the specimen resulting from postmortem formation.

In vitro studies have identified other volatile substances that may be produced during the decom-

position process. Volatiles that have been identified include acetone, isopropanol, n-butanol, t-butanol, isoamyl alcohol, and n-amyl alcohol. In a manner similar to ethanol formation, the specific volatile or volatiles produced is dependent on storage conditions.

See also: **Alcohol**: Blood; Body Fluids; Breath.

Further Reading

American Medical Association (1970) *Alcohol and the Impaired Driver*. Chicago: National Safety Council.

Borkenstein RF (1985) Historical perspective: North American traditional and experimental response. *Journal of Studies on Alcohol* suppl 10.

Corry JE (1978) Possible sources of ethanol ante- and post-mortem: its relationship to the biochemistry and microbiology of decomposition. *Journal of Applied Bacteriology* **44**:1–56.

Dubowski KM (1989) *Stages of Acute Alcoholic Influence/Intoxication*. Oklahoma City: University of Oklahoma College of Medicine.

Garriott JC (1996) *Medicolegal Aspects of Alcohol*, 3rd edn. Tuscon, AZ: Lawyers & Judges Publishing.

Havier RG (1991) The interpretation of ethanol with drugs. *Forensic Science Review* **3**:41–56.

Kunsman GW, Manno JE, Manno BR, Kunsman CM and Przekop MA (1992) The use of microcomputer-based psychomotor tests for the evaluation of benzodiazepine effects on human performance: a review with emphasis on temazepam. *British Journal of Clinical Pharmacology* **34**:289–301.

Kunsman GW, Manno JE, Przekop MA, Manno BR and Kunsman CM (1992) The effects of temazepam and ethanol on human psychomotor performance. *European Journal of Clinical Pharmacology* **43**:603–611.

Lamminpaa A (1995) Alcohol intoxication in childhood and adolescence. *Alcohol and Alcoholism* **30**:5–12.

Levine B and Smith ML (1990) Stability of drugs of abuse in biological specimens. *Forensic Science Review* **2**:147–157.

McKim WA (1991) *Drugs and Behavior: An Introduction to Behavioral Pharmacology*, 2nd edn. Englewood Cliffs, NJ: Prentice Hall.

NHTSA (1990) Briefing Paper. *Drug Evaluation and Classification Program* NHTSA, US Department of Transportation.

NHTSA (1993) *Student Manual. Drug Evaluation and Classification Training Program*. 'The Drug Recognition School', NHTSA, US Department of Transportation.

O'Hanlon JF and deGier JJ (1986) *Drugs and Driving*. Philadelphia: Taylor and Francis.

O'Neal CL and Poklis A (1996) Postmortem production of ethanol and factors that influence interpretation: a critical review. *American Journal of Forensic Medicine and Pathology* **17**:8–20.

Post Mortem

A W Jones, Department of Forensic Toxicology, University Hospital, Linköping, Sweden

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0406

Introduction

Alcohol (ethanol, CH_3CH_2OH) tops the list of toxic substances encountered in forensic toxicology for the simple reason that heavy drinking and drunkenness are incriminated in many fatal accidents, trauma deaths, suicides, crimes of violence and antisocial behavior in general. Reports from accident and emergency service departments worldwide provide ample evidence to support the negative impact of alcohol abuse and alcoholism in society. The impairment caused by overconsumption of alcoholic beverages explains many accidents in the home, in the workplace and on the roads. Accordingly, measuring and interpreting the concentrations of alcohol in blood and other biological specimens are routine procedures in forensic medicine and toxicology. The existence of threshold limits of blood-alcohol concentration (BAC) for driving a motor vehicle, such as 80 mg $100\,ml^{-1}$ ($0.80\,g\,l^{-1}$) in UK and 50 mg $100\,ml^{-1}$ ($0.50\,g\,l^{-1}$) in most of the other European nations, means that the results of forensic alcohol analysis have important social and medicolegal ramifications.

Qualitative and quantitative analysis of ethanol in postmortem specimens is a relatively simple analytical procedure and with gas chromatographic methods, accurate, precise, and specific results are possible. However, difficulties arise when the concentration of alcohol in a postmortem blood specimen is interpreted and conclusions are drawn about a person's state of inebriation at the time of death. Translating a BAC determined at autopsy (called necropsy in UK) into the amount of alcohol in the body at the time of death is subject to considerable uncertainty.

A major problem associated with postmortem alcohol analysis is the risk that the alcohol, at least in part, was generated or destroyed between the time of death and the time of the autopsy or after taking the specimens and performing the toxicological analysis. This becomes a major dilemma when decomposed bodies are examined and requests are made for alcohol analysis. The blood-glucose concentration increases after death owing to mobilization and hydrolysis of glycogen stores in the liver and muscle tissue thus providing abundant substrate for microbial synthesis of alcohol. This occurs by various processes when the conditions of time, temperature

and number and nature of available microorganisms are optimal.

After death the process of autolysis begins and body compartments are progressively destroyed. Within a few hours of death, bacteria and microorganisms begin to spread from the gastrointestinal canal through the portal system, and eventually reach deep into the vascular system. If alcohol was present in the stomach at the time of death, e.g. if the deceased had consumed alcohol before a fatal accident, this might diffuse into surrounding tissues such as the liver, heart, lungs and major blood vessels. For this reason the analysis of alcohol in stomach contents is a common practice in forensic toxicology to compare with blood-alcohol concentration thus providing a clue as to whether the person had died shortly after drinking alcohol.

Analysis of Alcohol in Postmortem Specimens

The methods suitable for analyzing alcohol in postmortem specimens are essentially the same as those used when specimens are taken from living subjects, e.g. drinking drivers or emergency service patients. The units used to report the concentration of alcohol determined in blood and other body fluids differ from country to country, e.g. $mg\,100\ ml^{-1}$ (UK), $g\,100\,ml^{-1}$ (USA), $g\,kg^{-1}$ or $mg\,g^{-1}$ (Scandinavia and Germany) and $g\,l^{-1}$ or $mg\,ml^{-1}$ (mid- and southern Europe). Examples of the kind of units used to report blood-alcohol concentrations for clinical and forensic purposes are given in **Table 1**.

Quantitative methods for the determination of alcohol in blood and urine have been available for more than 100 years. Although by modern standards the first efforts were rather primitive, the results at least took the guesswork out of deciding whether gross intoxication might have contributed as a cause of death. Ethanol had to be separated from the biological matrix by distillation or diffusion followed by oxidation with excess potassium dichromate in strong sulfuric acid and back titration of the amount of oxidizing agent remaining with sodium thiosulfate and iodometric titration. The major uncertainty with wet-chemical oxidation methods stemmed from the fact that other organic volatiles that might have been present in postmortem blood were oxidized along with ethanol leading to falsely high blood-alcohol concentrations being reported. The wet-chemistry oxidation procedures were replaced by milder and more selective enzymatic oxidation in the 1950s and under these conditions only a few other alcohols (propan-1-ol, isopropanol, butan-1-ol) represented any real interference problem.

In the 1960s gas chromatographic (GC) methods were developed for the analysis of alcohol in blood and urine and these have dominated ever since. GC methods had the distinct advantage of providing a qualitative screening analysis based on the retention time of the substance together with a quantitative analysis based on the detector response as reflected in peak area or peak height on the resulting chromatogram. The response of the flame ionization detector (FID) is remarkably linear over the range of concentrations ($0–800\,mg\,100\,ml^{-1}$) encountered in postmortem specimens. The limit of quantitation of alcohol in postmortem blood specimens under routine conditions is about $10\ mg\,100\ ml^{-1}$ ($0.1\ g\,l^{-1}$) and analytical results below this threshold are generally reported as negative in postmortem work.

Gas chromatography coupled with the headspace sampling technique (HS-GC) still remains the method of choice for forensic analysis of alcohol in specimens from living and dead persons. **Figure 1** gives a schematic representation of the headspace analysis procedure showing two chromatographic traces with propan-1-ol and t-butanol as internal standards. In postmortem toxicology two stationary phases (SP) are necessary to enhance specificity for ethanol. These are denoted as SP-1 and SP-2 in Fig. 1. When only a single stationary phase is used for GC-analysis,

Table 1 Concentration units used to report blood-alcohol concentration in clinical and forensic science.

mg ml^{-1} or *g l^{-1}*	*mg 100 ml^{-1}* or *mg dl^{-1}*	*g 100 ml^{-1}* or g%, w/v	*g kg^{-1}* or mg g^{-1a}	*mM* or *mmol l^{-1b}*
0.50	50	0.05	0.47	10.8
0.80	80	0.08	0.76	19.1
1.00	100	0.10	0.95	21.7
2.00	200	0.20	1.89	43.4
5.00	500	0.50	4.74	108.6

[a] The unit used in Scandinavia and Germany assuming a specific weight of 1.055 for whole blood, that is, 1 ml = 1.055 g.
[b] The SI unit used in clinical chemistry (mg ml $^{-1}$ × 1000)/46.05 where 46.05 is the molecular weight of ethanol.

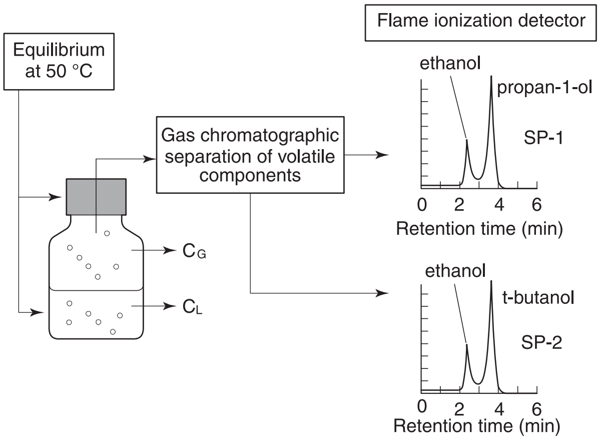


Figure 1 Schematic diagram showing analysis of alcohol in biological samples by headspace gas chromatography with two different stationary phases (SP1 and SP2) for packing the chromatographic columns and two different internal standards (propan-1-ol and t-butanol) for diluting the specimens prior to analysis.

there is always a risk that ethanol and some other volatile component in the blood, such as a product of putrefaction or fermentation, might have the same retention time. Making a duplicate determination of the blood-alcohol concentration with two different column packing materials for the chromatography gives two different retention times for ethanol which minimizes or eliminates the risk of obtaining coincident retention times.

Examples of the retention time of ethanol compared with other low-molecular-weight volatiles with two widely used stationary phases for GC analysis are given in **Table 2**. If needed, the chromatographic operating conditions (e.g. oven temperature, carrier gas flow) can be optimized to give better separation of peaks for the various substances. The use of two chromatographic systems is a mandatory requirement in postmortem toxicology when ethanol is the substance analyzed. Packed columns made from glass or stainless-steel tubes which are 2 m long and 3 mm internal diameter are still widely used for forensic alcohol analysis. However, major developments have occurred in separation science for gas chromatography and today wide-bore capillary columns dedicated for blood-alcohol analysis are available such as Rtx-BAC1 and Rtx-BAC2. Whatever the chromatographic method used, a minimum requirement must be to determine methanol, ethanol, acetone and propan-2-ol at the same time in a single run. These are the most frequently encountered low-molecular-weight volatile substances in postmortem blood specimens.

The precision of HS-GC methods of blood-alcohol analysis expressed as coefficient of variation (CV) is often less than 1% within a single laboratory. The variation between laboratories, reflecting the reproducibility of the method, when aliquots of the same blood specimen are sent to specialist forensic laboratories, is about 3% CV, and in hospital clinical laboratories CVs of 6–8% are reported. In postmortem toxicology, the sampling variation and magnitude of site-to-site differences in blood alcohol concentration often exceeds these pure analytical variations.

Another way of enhancing selectivity for identification of ethanol is to make use of an independent analytical principle such as enzymatic or chemical oxidation together with the usual gas chromatographic analysis. However, the current trend is

Table 2 Retention times (RT) of ethanol and other low molecular volatile substances analyzed by headspace gas chromatography on two different stationary phases commonly used in forensic toxicology laboratories. RTs relative to propan-1-ol as internal standard are shown in brackets

Substance	*Retention time (min) Carbopak C as column packing material*[a]	*Retention time (min) Carbopak B as column packing material*[b]
Acetaldehyde	0.56 (0.38)	0.53 (0.29)
Acetone	1.00 (0.68)	0.86 (0.46)
Butan-1-ol	4.68 (3.16)	4.11 (2.22)
Butan-2-ol	2.99 (2.02)	2.53 (1.36)
Ethanol	0.72 (0.49)	0.98 (0.53)
Methanol	0.49 (0.33)	0.67 (0.36)
Methyl ethyl ketone	2.45 (1.66)	1.49 (0.81)
Isopropanol	1.16 (0.78)	1.31 (0.71)
Propan-1-ol	1.48 (1.00)	1.85 (1.00)
t-butanol	1.90 (1.28)	1.68 (0.91)

[a] Carbopak C, 0.2% Carbowax (polyethylene glycol) 1500.
[b] Carbopak B, 5% Carbowax (polyetheylene glycol) 20M.

towards the use of mass selective detectors for verification of ethanol in blood specimens. The electron-impact mass spectrum of ethanol shows prominent ion-fragments with a base peak at m/z 31 (common for primary alcohols), m/z 46 (molecular ion) and also at m/z 45 (molecular ion -1). Compact bench-top GC-MS instruments are now affordable and these are widely available in forensic toxicology laboratories. These instruments offer excellent opportunities for unequivocal identification of alcohols and other volatiles in blood and tissue specimens obtained at postmortem as well as in the poisoned patient.

Sometimes a rapid assessment of whether alcohol intoxication was a contributing factor in a person's death is considered necessary. For this purpose, several on-the-spot methods of alcohol analysis are available. One such method consists of a disposable dip-stick device, which was originally developed for the analysis of alcohol in saliva specimens from living subjects. This saliva alcohol test is called QED™ and stands for quantitative enzyme diagnostics. This device can be used to measure alcohol in 'clean' biological liquids such as vitreous humor or cerebrospinal fluid (CSF) or in blood or plasma after precipitation of proteins. The principle of the QED™ analysis involves enzymatic oxidation of ethanol by alcohol dehydrogenase (ADH) and the coenzyme NAD^+. The endpoint is coupled to a blue-color reaction. The result of the test is obtained after about 1 min by reading a thermometer-like scale. The length of the blue stain in the capillary tube is proportional to the concentration of alcohol in the sample analyzed.

Alternatively, a handheld instrument (Alcolmeter S-D2) intended primarily for breath-alcohol testing can be modified for headspace sampling above liquids. The Alcolmeter incorporates an electrochemical sensor for the oxidation of alcohol and about 0.50 ml of headspace vapor is aspirated from above the liquid tested (e.g. blood or urine) equilibrated in an airtight container held at constant temperature. The result of the analysis is obtained in about 30 s after sampling. Although the results obtained with these quick and easy methods are less reliable than those obtained by gas chromatography, they at least indicate whether the person might have been under the influence of alcohol at the time of death.

The biological specimens received by the laboratory should be allowed to acclimatize to room temperature before aliquots are removed for analysis. Blood samples should be made homogenous and if necessary any clots homogenized to obtain a uniform sample. Otherwise, the clot can be separated by centrifugation and the clear supernatant analyzed remembering to take into account the water content of the specimen which might be different from that of whole blood. Urine, vitreous humor and CSF are usually analyzed directly without any pretreatment.

Fate of Alcohol in the Body

Traces of alcohol occur naturally in body fluids as products of metabolism and also by the microbial fermentation of sugars in the gut. However, this endogenously produced ethanol lacks any forensic significance because the concentration in peripheral venous blood from abstaining subjects determined by headspace gas chromatography are below $0.001\,g\,l^{-1}$ ($0.1\,mg\,100\,ml^{-1}$). Even in subjects suffering from various metabolic diseases such as diabetes mellitus, cirrhosis, hyperthyroidism etc., the concentrations of endogenous ethanol are similar to values observed in healthy individuals.

After drinking alcoholic beverages, the alcohol (ethanol) contained in beer, wine or spirits is diluted with the stomach contents before being absorbed and transported by the blood to all body organs and tissues. Alcohol distributes throughout the total body water without binding to plasma proteins. The solubility of ethanol in fat and bone is negligible. How fast alcohol enters the bloodstream depends on many variable factors, especially the speed of gastric emptying as controlled by the pyloric sphincter. Alcohol absorption is relatively slow through the stomach mucosa, which is less permeable to small molecules than the duodenal or jejunal mucosa. Also, the much larger absorption surface area available in the upper part of the small intestine facilitates rapid absorption of alcohol, which requires no prior digestion. Factors that delay gastric emptying such as the presence of food in the stomach before drinking, various medications, smoking, blood-sugar level and the time of day will impact on the rate of absorption of ethanol and influence the blood-alcohol concentration reached.

After drinking small doses of alcohol (one to two drinks) some of the alcohol might become metabolized in the stomach mucosa or during the first passage of the blood through the liver. The presystemic breakdown of alcohol in the gut or the liver before reaching the systemic circulation is referred to as 'first-pass metabolism'. The mucous membranes of the stomach contain the enzyme ADH although in much smaller amounts than in the liver where the bulk of the metabolism of ethanol occurs. Gastric ADH differs from hepatic ADH in other respects such as the optimal K_m and V_{max} values for oxidation of alcohol.

The blood-alcohol concentration reached after drinking alcoholic beverages depends on the amount consumed (dose), the speed of drinking, the rate of absorption from the gut and also on the person's body weight, age and gender. Having a higher proportion of fat gives a higher BAC for the same dose of alcohol consumed because leaner individuals have more body water into which the alcohol can be diluted. Since women tend to be smaller than men and also have more fatty tissue and less body water, a given amount of alcohol in a female drinker yields a higher BAC and a correspondingly greater effect on the brain (impairment) and more damage to organs and tissues. Likewise in older individuals, who generally have more fat per kg body mass than younger people, studies have showed that higher BACs are reached in the aged.

Once absorbed from the gut, alcohol is transported via the portal vein to the liver where enzymes begin the process of breaking down the alcohol to clear it from the bloodstream. The principal alcohol-metabolizing enzyme is class I hepatic alcohol dehydrogenase (ADH), which converts ethanol into a toxic metabolite, acetaldehyde. Fortunately, this noxious substance is swiftly transformed into acetate by another hepatic enzyme called aldehyde dehydrogenase (ALDH). In this two-stage biotransformation process, the coenzyme nicotinamide adenine dinucleotide (NAD^+) is reduced to NADH. The excess NADH in the hepatocytes during ethanol oxidation disrupts many of the normal metabolic processes that require the same coenzyme (NAD^+). This leads, among other things, to reduced synthesis of glucose, increased concentration of lactate, altered fatty acid metabolism and accumulation of fat in the liver as some of the consequences. The bulk of the dose of ethanol (95–98%) is oxidized to CO_2 and H_2O, and the remainder (2–5%) is excreted, unchanged, in sweat, urine and expired air. A very small fraction of the alcohol ingested is converted by the liver into ethyl glucuronide and this water-soluble minor metabolite can be measured in urine specimens to confirm that a person had actually taken alcohol.

Distribution of Alcohol in Body Fluids and Tissue

Alcohol and water mix together in all proportions and only a very small fraction of the total amount of alcohol absorbed into the blood penetrates into fatty tissue and bone. Accordingly, the distribution of alcohol in body fluids and tissue after diffusion equilibrium is complete follows the distribution of water in the body. Urine, vitreous humor (VH), and CSF, which consist of 98–99% w/w water, can therefore be expected to have higher concentrations of alcohol than whole blood, which is 80% w/w water on the average. Organs and tissue such as skeletal muscle, brain, liver and kidney which contain somewhat less water (75–78% w/w) will accordingly contain lower concentrations of alcohol compared with the same mass of blood. Moreover, liver and kidney have considerable enzymatic activity and the concentration of alcohol in these organs decreases for various periods of time after death owing to on-going metabolic processes.

Table 3 gives the water contents of body fluids and tissues taken at postmortem and used for analysis of alcohol. These values provide a rough estimate of the relative concentration of alcohol expected in various biofluids and tissues because the ratios of the water content should correspond to the distribution ratios of alcohol provided that diffusion equilibrium of alcohol is complete. The stage of alcohol absorption and distribution in the body and the pooling of urine in the bladder during which time the blood-alcohol concentration is changing are important to consider when urine/blood ratios and CSF/blood ratios of alcohol are interpreted.

After absorption and distribution of alcohol are complete and the concentration in blood begins to decrease, the organs and tissue such as skeletal muscles return alcohol into the venous blood and periph-

Table 3 Average water-content of body fluids, organs and tissue in relation to the expected concentrations of alcohol. Note that the result from analyzing an alternative biological specimen should not be used to estimate blood-alcohol concentration in any individual case

Biological specimen	*Water content (% w/w)*	*Alcohol ratio relative to whole blood*[a]
Whole blood	80	1.0
Plasma/serum	92	1.1–1.2
Erythrocytes	70	0.8–0.9
Urine	98–99	1.2–1.4[b]
Vitreous humor	99	1.1–1.3
Cerebrospinal fluid	98–99	1.1–1.3[c]
Bile	87–97	0.8–1.0
Synovial fluid	92–96	1.1–1.2
Liver	80	0.6–0.8
Brain	75	0.8–1.0
Skeletal muscle	76	0.8–0.9
Kidney	79	0.6–0.7

[a] The ratios of alcohol concentration differ widely in any individual case depending on the concentration of alcohol present and the time after drinking when death occurred.
[b] Average value for specimens taken at autopsy; note that the ratio depends to a great extent on the stage of alcohol absorption and distribution in the body.
[c] Lumbar fluid taken during postabsorptive phase.

eral circulation. The venous blood therefore contains a somewhat higher concentration of alcohol than arterial blood when the blood-alcohol curve enters the postabsorptive stage. However, these arterial–venous differences in alcohol concentration lack significance in postmortem work.

Toxicity of Alcohol

Fatalities caused by abuse of alcohol and drunkenness occur daily throughout the world and heavy drinkers and alcoholics are over-represented among individuals committing suicide. When sudden unnatural deaths are investigated, especially among males, the BACs at autopsy together with other information corroborate recent heavy drinking. Elevated blood-alcohol concentrations are the most common finding in medicolegal autopsies often in combination with narcotics or prescription drugs. Indeed, alcohol intoxication or drug–alcohol interactions should not be overlooked when sudden natural deaths are investigated and screening of all out-of-hospital deaths for elevated blood-alcohol concentration has been suggested.

Deaths attributable to alcohol impairment include all kinds of accidents and especially road-traffic fatalities, which are the leading cause of death among people aged under 35 years for drivers, passengers and pedestrians. Statistics indicate that in approximately 30–40% of fatal road-traffic accidents, especially single vehicle accidents, the driver had been drinking, and that blood-alcohol concentrations at autopsy often exceeded 150 mg 100 ml^{-1} (1.5 g l^{-1}).

The acute toxicity of alcohol is well documented for inexperienced drinkers who consume too much too quickly, leading to gross impairment and alcoholic coma. Alcohol kills by its direct toxicological effect on the brain, e.g. depression of the respiratory center in the medulla, with paralysis of respiration. Under these circumstances the BACs at autopsy are often 300–400 mg 100 ml^{-1} (3–4 g l^{-1}), or higher. In combination with other sedative drugs (e.g. barbiturates and benzodiazepines), for which the pharmacological effects are additive, death may result after drinking much smaller amounts of alcohol. Long-term abuse of alcohol eventually leads to malfunctioning of body organs and tissue necrosis. Death from liver cirrhosis is a fairly common finding in chronic alcoholics. Secondary causes of death include aspiration of vomit and suffocation while in a deep comatose drunken stupor. This mode of death is well documented in teenagers and young people unaccustomed to heavy drinking and for whatever reason have engaged in forced consumption of alcohol, such as occurs during drinking competitions.

Some people consume large quantities of alcohol daily and build-up a pronounced central nervous system tolerance, which might explain the very high blood-alcohol concentration reported in autopsy material (4–5 g l^{-1}). Death from depression of the respiratory center in the brain or a combination of gross intoxication and exposure, e.g. hypothermia is not uncommon among skid-row alcoholics. Chronic intake of alcohol results in metabolic disturbances such as hypoglycemia, lactic acidosis and ketoacidosis. These conditions are often exaggerated by poor dietary habits sometimes with fatal outcome.

Many attempts have been made to correlate blood-alcohol concentration with degree of impairment, but the association is weak and only general guidelines can be given, in part owing to the development of functional tolerance. The blood-alcohol concentration necessary to cause death varies widely but is often quoted as being 350–450 mg dl^{-1} although many exceptions exist and some people attempt to drive a motor vehicle after drinking to reach these levels.

Sampling Considerations

Obtaining representative samples is a critical element in all analytical work. The late W.J. Youden of the National Bureau of Standards, Washington, DC is quoted as saying: 'It is common knowledge that the result of an analysis is no better than the sample used for analysis. The qualifying phrase on the sample as received should be included in a report.'

Analytical chemists and the methods they use are carefully scrutinized regarding precision, accuracy and selectivity of the results, but less attention is usually given to the way the sample was obtained and its condition on arrival at the laboratory. Indeed, the variance and uncertainty associated with sampling often dominates the total variability of the analytical method. For this reason, autopsy specimens should be taken in accordance with a standard protocol and all materials carefully labelled including identification of the deceased, the date and time of autopsy and the type and amount of specimen collected. These aspects of the sampling procedure become important if a second autopsy has to be performed, and new specimens are taken for toxicological analysis. The size, shape, and composition of the containers used to hold the body fluids should be appropriate for the kind of specimen collected. The containers (previously unused) are generally made of glass or polypropylene

with tight fitting screw tops. This reduces the likelihood of contamination of specimens and losses of sample occurring owing to a leaky container or adsorption or reaction of the analyte with the container material. The mode of transport, storage and overall security of the materials sent for analysis often needs to be documented in medicolegal casework whenever the results of alcohol analysis are challenged in court.

The cleanliness of the sampling equipment (sterile syringes and needles) and any transfer lines also warrant attention to avoid contamination of specimens with extraneous solvents used in the autopsy room. For volatile substances, like ethanol, use of airtight containers with small air-spaces helps to minimize losses by evaporation. Whenever possible the biological specimens should be stored in the cold (4°C) before shipment to the laboratory. Finally a system should be devised to permit the definitive identification of samples received by the laboratory including the kind of specimen, the site of sampling blood, the volume of specimen obtained, and not least the person responsible for inspecting the specimen. All this falls under the rubric 'chain-of-custody' which becomes critical whenever the integrity of the sample has to be defended in court.

Examples of the biological specimens taken at autopsy and subsequently used for alcohol analysis are presented in **Table 4**. The rank order of preference is femoral venous blood, bladder urine or vitreous humor, and cerebrospinal fluid or various combinations of these depending on availability, e.g. blood and urine or blood and vitreous humor.

Blood Samples

The concentration of alcohol in a sample of postmortem blood can provide useful information about the BAC at the time of death and within limits the amount of alcohol the person might have consumed.

Table 4 Biological specimens available for analysis of alcohol and sometimes taken during postmortem examination

• Whole blood	• Bile
Heart (cardiac)	• Stomach contents
Cubital vein	• Synovial fluid
Jugular vein	• Bone marrow
Femoral vein[a]	• Organs and tissue
Intercranial blood	Skeletal muscle
• Urine[a]	Brain
• Vitreous humor[a]	Liver
• Cerebrospinal fluid	Kidney

[a] Preferred specimens

The effects of alcohol and other psychoactive substances on performance and behavior tend to be correlated with the concentrations infiltrating the brain and the person's BAC provides the best indirect estimate of central nervous system (CNS) exposure to the drug. Tabulations of therapeutic, toxic and fatal concentrations of drugs are available for comparison to help with interpretation of the concentrations determined at autopsy.

The kind and quality of the blood sample available at autopsy depends to a large extent on the condition of the corpse and in particular the existence of severe trauma and whether any evidence of decomposition exists. Taking relevant samples is fundamental to allow correct interpretation of the analytical results. Blood ($\sim$50 ml) should be taken from an extremity, such as an undamaged femoral vein in the leg or a jugular vein in the neck and the specimens should be obtained early during the autopsy and before evisceration. A recommended practice is to clamp the femoral vein and then withdraw a suitable blood sample with a sterile syringe and wide bore needle. Blood samples intended for alcohol analysis should not be taken from the pericardium, abdominal or thoracic cavities because at these anatomical locations there is an increased risk of contamination by alcohol spreading from the gut. Although heart blood is occasionally used as a specimen for postmortem alcohol analysis, this is not recommended owing to the risk of contamination with alcohol diffusing from the stomach or upper airways.

Blood specimens submitted for alcohol analysis are often clotted and completely hemolyzed, and occasionally, also diluted with other biological fluids. Making a blind-stick through the chest enhances the risk of contamination of samples with ethanol that might have diffused from the stomach into the pleural and pericardial spaces. This problem is especially acute if the stomach has ruptured as often happens in multitrauma deaths, e.g. fall from a high building, motor-vehicle accident or plane crash. The postmortem diffusion of alcohol and other drugs is a recurring issue in postmortem toxicology when the analytical results are interpreted. The lungs, cardiac blood and abdominal spaces might also be contaminated with alcohol if the deceased aspirated vomit when high concentrations of alcohol remained in the stomach.

The sampling tubes and containers used to collect and store postmortem specimens for alcohol analysis must contain sufficient sodium or potassium fluoride so that the final concentration of preservative is approximately 2%, w/w. The fluoride ion is a potent enzyme inhibitor and helps to prevent glycolysis and any production of alcohol by fermentation

if viable yeasts or other microorganisms are available to ferment glucose to alcohol.

Obtaining intracranial blood from a subdural or subarachnoid hematoma is one useful strategy to investigate whether the person had been drinking alcohol before receiving a blow to the head, fracture of the skull and cerebral hemorrhage. If the victim survives several hours after the trauma the concentration of alcohol in peripheral venous blood might have decreased to zero owing to metabolism occurring in the liver. Because circulation in a cerebral blood clot is diminished or nonexistent the concentration of alcohol present should reflect the concentration in peripheral blood at an earlier point in time. Thus, by comparing the alcohol concentration in subdural blood with, for example, femoral venous blood gives a clue about the person's BAC at the time of the accident. However, low concentrations of alcohol in intracranial blood clots might have been produced by microbial activity, which underscores the need for obtaining supporting evidence of alcohol consumption such as the analysis of urine or vitreous humor. It is not easy to introduce fluoride ions into a blood clot.

If necessary, the blood-alcohol concentration determined at autopsy can be translated into the amount of alcohol in the body at the time of death. With additional information (e.g. gender, body weight, time of starting to drink) the total quantity of alcohol consumed can also be estimated although the results are subject to wide variations. The scientific basis for making these blood-alcohol calculations, which almost always involves use of the Widmark equation, is presented elsewhere.

Vitreous Humor

Vitreous humor or fluid is the transparent gelatinous material filling the eyeball just behind the lens. This watery fluid makes an ideal specimen for forensic analysis of alcohol because of the isolated location of the sampling site, that is, the remoteness of the eyes from the gut, thus minimizing the risk of contamination with microorganisms or diffusion of alcohol from the stomach. The sampling and analysis of alcohol in vitreous humor (VH) is therefore highly recommended as a complement to taking blood samples for toxicological analysis and specimens of VH can be obtained without making a full autopsy. Comparing the concentration of alcohol in VH with the blood-alcohol concentration allows a check on whether postmortem synthesis of alcohol in the blood samples needs to be considered. Good agreement has been observed for the concentrations of alcohol determined in VH retrieved from both eyeballs.

Experience has shown that VH is more resistant to putrefactive changes than peripheral blood samples especially in highly traumatic deaths, e.g. aircraft accidents. When there is extensive trauma to the body, the spread of bacteria from the alimentary canal to various parts of the vascular system is much more likely. Under these circumstances, sampling and analysis of VH becomes virtually essential to allow making a reliable interpretation of the prevailing blood-alcohol concentration at the time of death. Moreover, it remains feasible to sample VH for analysis of alcohol when the corpse has become moderately decomposed. Finding a negative concentration of alcohol in VH and an elevated BAC strongly suggests that alcohol has been produced in the blood after death.

The concentration of alcohol in VH should exceed that of the femoral venous blood as there is roughly 10–20% more water in the eye fluid. The VH/BAC ratio depends to some extent on the time after drinking when death ensued, that is, on the stage of absorption and distribution of alcohol in the body. During or shortly after the end of drinking, one might expect the VH/BAC ratio of alcohol to be less than or close to unity whereas in the postabsorptive stage of alcohol pharmacokinetics when equilibration of alcohol in all body fluids is complete, the VH/BAC ratio should be about 1.2:1.

Figure 2 shows a scatter plot of the concentrations of alcohol in VH and in femoral venous blood in

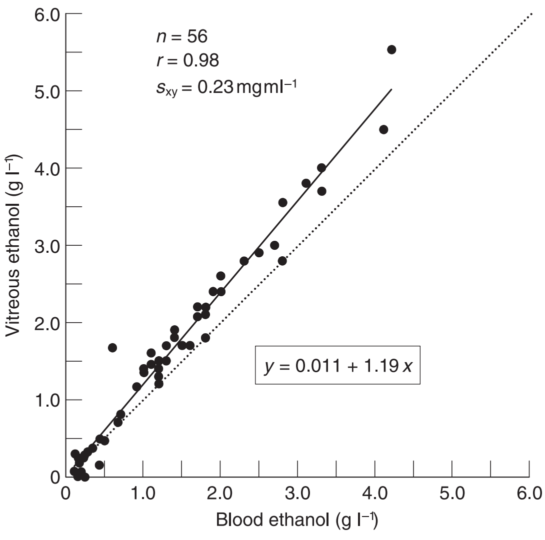


Figure 2 Scatter plot showing the relationship between the concentration of alcohol in femoral venous blood and vitreous humor in 56 autopsy cases (A.W. Jones, unpublished work).

samples from 56 autopsies. The correlation coefficient was high ($r = 0.98$) although the scatter of the points around the regression line as reflected by the standard error estimate (s_{yx}) was large, being $0.23\,mg\,l^{-1}$ so that 95% of cases should be expected to fall within $\pm\ 0.46\,mg\,l^{-1}$ ($2 \times s_{yx}$). It is obvious, therefore, that estimating blood-alcohol concentration indirectly from analysis of VH is rather uncertain if 95% limits of agreement are used, and even more uncertain if 99% confidence is required. The negligible intercept ($0.01\,g\,l^{-1}$) indicates rapid equilibration of alcohol between blood and eye fluids with no pooling of the alcohol and a minimal concentration–time lag. The regression coefficient of 1.19 indicates that VH-alcohol is 19% higher than the corresponding blood alcohol concentration in this material.

The proper role of VH as a biological specimen for alcohol analysis in postmortem toxicology is to compare results with the BAC and thus to corroborate the presence of alcohol in the body at the time of death. Without the analysis of VH, the blood-alcohol concentration alone, especially if this is low ($< 0.5\,g\,l^{-1}$ or $< 50\,mg\ 100\,ml^{-1}$), makes the conclusion that a person was drinking before death highly contentious. VH is also the most useful specimen for alcohol analysis whenever embalmed bodies are examined. Embalming fluids contain, among other things, formalin and also various alcohols (mainly methanol).

Urine Samples

The sampling and analysis of bladder urine aids in the interpreting of BAC determined at autopsy in the same way as VH. The total volume of urine in the bladder should be measured and an aliquot (10–50 ml) taken for toxicological analysis. The container into which the urine specimen is placed should be prepared in advance with sodium or potassium fluoride so that the final concentration is about 2%, w/v, taking care that the salt dissolves in the urine by mixing thoroughly. Sodium fluoride is less soluble in water than potassium fluoride.

Drugs and their metabolites are excreted in the urine and are present in higher concentrations than in the peripheral blood. The normal rate of urine production is $1\,ml\,min^{-1}$ ($60\,ml\,h^{-1}$), but diuresis is enhanced after drinking alcohol especially when the BAC curve is rising. Studies have shown that the mean urine/blood ratio of alcohol increases from 1.25:1 to 1.65:1 as the volume of urine in the bladder at autopsy increases from < 5 ml to 200–400 ml. Higher urine-to-blood alcohol concentration (UAC/BAC) ratios tend to be associated with low concentrations of alcohol, because like percentages the value of a ratio depends on the size of the denominator. At low BAC ($< 0.5\,g\,l^{-1}$ or $50\,mg\ 100\,ml^{-1}$), abnormally high UAC/BAC ratios are generally obtained, in part because of the properties of ratio variables.

The UAC/BAC ratio might give a hint about the status of alcohol absorption at the time of death, which could have forensic significance when fatal accidents are investigated. **Figure 3** shows a plot of urine alcohol and blood alcohol in 21 subjects who first emptied their bladders and then drank a moderate dose of alcohol in the morning without having eaten breakfast. Note that UAC is less than BAC during the first 1–2 h after drinking and higher than BAC during the remainder of the postabsorptive phase 2–7 h after drinking. Finding a UAC/BAC ratio of 1.1 or less suggests that absorption and distribution of alcohol was incomplete and that the person might therefore have consumed alcohol within 2 h of death. In contrast, a UAC/BAC ratio of 1.3 or more indicates that the alcohol was already equilibrated in body fluids and tissues and that consumption had probably ended several hours before death. Abnormally high or low UAC/BAC ratios are sometimes encountered in postmortem work for the reasons given in **Table 5**.

Figure 4 shows a scatter plot of UAC and BAC in 56 postmortem samples. The individual values are highly correlated ($r = 0.94$) but the scatter around the regression line was considerable ($s_{yx} = 0.48$). The individual UAC–BAC fluctuations were so wide that attempts to estimate BAC indirectly from analysis of UAC cannot be recommended. At zero BAC the mean UAC (y intercept) was $0.25\,g\,l^{-1}$, which indicates a pooling of urine in the bladder during which time the BAC continues to decrease owing to metabolism of alcohol in the liver. Besides, this constant bias of $0.25\,g\,l^{-1}$, the UAC showed a proportional

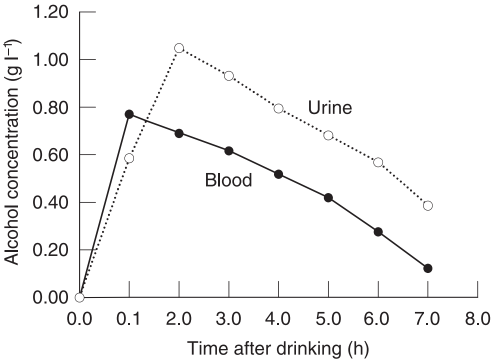


Figure 3 Mean concentration–time profiles of alcohol in urine and blood after 21 subjects drank a moderate dose of alcohol on an empty stomach (A.W. Jones, unpublished work).

Table 5 Interpreting urine/blood ratios of alcohol in autopsy specimens

- Appreciable concentration of alcohol in urine with low or zero concentration in blood ($<0.1\,g\,l^{-1}$ or $<10\,mg\,100\,ml^{-1}$). This suggests elimination of alcohol from the bloodstream owing to metabolism with a subsequent pooling of urine in the bladder. These situations are encountered when death is caused by a blow to the head resulting in cranial fracture, cerebral hemorrhage, and when the victim survived the traumatic injuries for several hours.
- Another explanation for elevated UAC and low or zero BAC might be if alcohol was synthesized in the urine owing to infection with bacteria and/or yeast and if the specimen contained sugar, e.g. in a person with diabetes.
- Zero or very low concentration of ethanol in urine ($<0.1\,g\,l^{-1}$ or $<10\,mg\,100\,ml^{-1}$) and appreciable concentration of alcohol in blood samples. These situations suggest microbial synthesis of ethanol in blood, which is often seen after traumatic injuries and decomposition of the body. An alternative explanation is death occurring immediately after drinking a large amount of alcohol with the bladder containing an alcohol-free pool of urine.
- Abnormally high urine/blood ratios of alcohol are seen when low concentrations of alcohol are present in body fluids or when the bladder contains unusually large volumes of urine ($>300\,ml$).

bias of 15% compared with BAC as indicated by the regression coefficient of 1.15.

The main advantage of urine over blood is that invasion of the bladder by microbes and yeasts appears to be less of a problem and except in those cases with a history of diabetes, urine does not normally contain any significant amounts of sugar for alcohol fermentation. The combination of sugar, viable yeast or other microorganisms, and optimal time and temperature conditions leads to the formation of alcohol in urine within about 24–36 h after death; one molecule of glucose produces two molecules of ethanol and carbon dioxide during the fermentation process. A glycosuria of $500\,mg\,100\,ml^{-1}$ could result in a UAC of $250\,mg\,100\,ml^{-1}$ on complete fermentation by a yeast such as *Candida albicans*. The concentration of alcohol measured in urine should not be used to estimate the concentration of alcohol in the body at the time of death.

In aviation fatalities where polytraumatic deaths are the rule rather than the exception, drawing correct conclusions about alcohol consumption from analysis of a single specimen of blood or tissue is very difficult. Rupturing of the stomach and bursting of the bladder are commonly seen in victims of plane crashes making it difficult to obtain urine for toxicological analysis. In aircraft accidents, there is always a bigger risk for postmortem synthesis of alcohol because of the abdominal trauma and microorganisms from the gut spread more quickly throughout the body. **Table 6** lists some of the problems associated

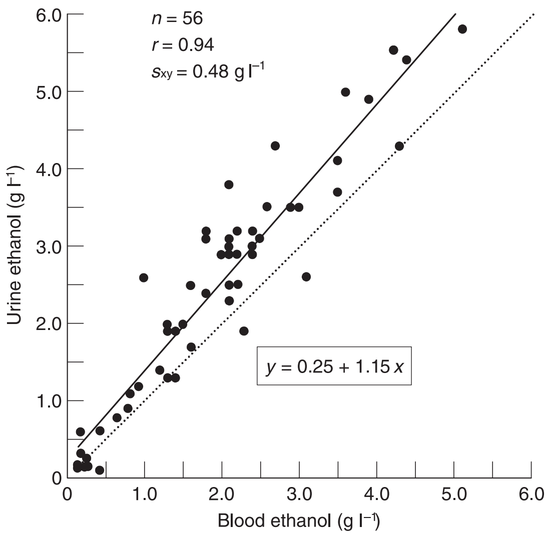


Figure 4 Scatter plot showing the relationships between the concentration of alcohol in femoral venous blood and bladder urine in 56 autopsy cases (A.W. Jones, unpublished work).

Table 6 Some factors that need to be considered when interpreting results of alcohol analysis in postmortem specimens

- Specificity of the analytical method used and whether other volatile substances could interfere with the analysis of alcohol.
- Water content of the biological specimens analyzed.
- Variations in alcohol concentration depending on sampling site in the vascular system.
- Variations in alcohol concentration depending on stage of alcohol pharmacokinetics when death occurred.
- Postmortem diffusion of alcohol from the stomach.
- Postmortem synthesis of alcohol by the action of a host of possible microorganisms acting on glucose, amino acids and other substrates.
- Postmortem losses by evaporation or degradation of ethanol by various microorganisms.
- Contamination of specimens with extraneous solvents during emergency service treatment in the hospital or the mortuary or the laboratory.
- Embalmed bodies and the kind of embalming fluid used and whether any alcohol present might have contaminated the material analyzed?

with the analysis and interpretation of alcohol concentrations in postmortem specimens.

Cerebrospinal Fluid (CSF)

The CSF is mainly composed of water (97–99%) and the concentration of alcohol reaching CSF indicates the concentration that has passed through the brain giving an indication of pharmacological exposure. CSF is obtained either by lumbar puncture at the base of the spine with a hypodermic syringe and wide-gauge needle or by withdrawing cisternal fluid by puncturing the back of the neck. CSF is a clean biological specimen well suited for the analysis of alcohol and other drugs. The CSF is relatively protected from fermentation processes and is sufficiently isolated so that diffusion of alcohol from the stomach is not a major problem. However, CSF might be contaminated with blood when death is caused by a blow to the head.

Figure 5 shows an example of the pharmacokinetic profiles of alcohol in venous blood and CSF obtained by lumbar puncture at various time intervals after subjects drank a moderate dose of alcohol in 5 min. The concentration–time profiles of alcohol in blood and CSF are shifted in time. This is particularly apparent for CSF drawn from the lumbar region but less so for cisternal fluid. This creates a pooling of alcohol in the lumbar fluid so the concentration of alcohol in lumbar CSF does not reflect the BAC at the time of death.

The concentration of alcohol in CSF clearly lags behind the BAC during absorption and distribution of alcohol and the CSF/BAC ratios are less than unity. After 60–90 min when BAC is decreasing, the CSF/BAC ratios are now 1.2–1.35 times the coexisting venous BAC. The CSF/BAC ratio of alcohol should be about 1.2:1 on the basis of differences in water content and complete equilibration in the body. In practice, the CSF/BAC ratio for lumbar fluid often exceeds 1.3:1 or 1.4:1 in the postabsorptive phase owing to the time lag and the fact that BAC is continuously decreasing. Note that the time lag for cisternal fluid is less than for lumbar fluid because of the slow movement of CSF down the spinal subarachnoid space. The concentration–time course of cisternal CSF therefore runs closer to BAC and the CSF/BAC ratio in the postabsorptive phase is accordingly less than for lumbar CSF fluid.

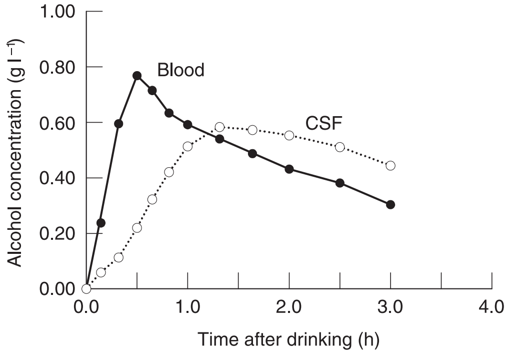


Figure 5 Mean concentration–time profiles of ethanol in cerebrospinal fluid (CSF) obtained by lumbar puncture and venous blood in nine subjects after ingestion of a moderate dose of alcohol (graph redrawn from Fleming and Stotz 1935).

Other Specimens

The first choice of body fluids for postmortem alcohol analysis are femoral venous blood, bladder urine and vitreous humor (Table 4). When these are not available other biological specimens or tissues are desirable and occasionally submitted for analysis of alcohol. These might include bile from the gall bladder, synovial fluid from the knee joint, marrow from the bones, as well as stomach contents. Tissue such as liver, brain, kidney and skeletal muscle have also served as specimens for analysis of alcohol in postmortem toxicology. In tissues capable of metabolizing ethanol, e.g. liver and kidney, a progressive decrease in the concentration of alcohol has been observed for various periods of time after death. The presence of oxygen in any surviving tissue and availability of NAD^+ are sufficient conditions for continued enzymatic oxidation.

Skeletal muscle contains glycogen which is converted into glucose after death providing an abundance of substrate for microbial synthesis of alcohol. Several reports indicate that alcohol is generated in muscle tissue within the first few days after death or during the time specimens are sent by mail to the laboratory for analysis. Alcohol is not evenly distributed in the brain owing to an uneven distribution of water so the results depend on which brain region was analyzed. In forensic practice, a useful strategy when dealing with tissue samples would be to prepare a homogenate of the material intended for analysis immediately after completion of the autopsy and making sure to add about 2% NaF as a preservative before sending the specimens for toxicological analysis.

It is important to consider the water content of organs and tissue when the results from analyzing unconventional specimens are interpreted. Water content is easily determined by drying a sample of tissue or fluid to reach a constant weight. Studies have shown that the water content of fresh whole blood is about $80 \pm 5\%$, w/w, but in postmortem blood the

water content is much more variable and tends to decrease with increasing time after death. Postmortem blood has a greater proportion of erythrocytes compared with fresh blood and the amount of water in postmortem specimens might vary from 60% to 90%, w/w, depending on various circumstances. Some investigators have recommended that blood and tissue water is determined by freeze drying or desiccation along with the blood-ethanol concentration. A correction could then be applied to the measured BAC for the diminished amount of water in postmortem specimens. Making these allowances for water-content have been advocated and applied by forensic pathologists in Germany when results of postmortem alcohol analysis are interpreted. However, there are so many other uncertainties when interpreting results of postmortem alcohol analysis that making a correction for water content might not be so important.

Stability of Alcohol in Postmortem Blood

The stability of alcohol in blood after death is another problem faced when the analytical results are evaluated and a judgment is made about a person's state of inebriation at the time of death. Comparing alcohol concentration in venous blood obtained shortly after death, e.g. at the scene of a road traffic accident with femoral venous BAC obtained at autopsy has shown that both increases and decreases in the concentration of alcohol can occur. However, the results of these studies are confounded by inherent site-to-site variations in concentrations of alcohol, any life-saving medication administered and the possibility of postmortem diffusion taking place prior to autopsy.

The mechanism of alcohol loss from tissues and body fluids after death might involve evaporation, enzymatic breakdown or microbiological degradation. A host of different microorganisms can utilize alcohol as a substrate. Loss of alcohol can occur in the interval between death and performing the postmortem examination, and from the time of autopsy to making the toxicological analysis. Whenever possible, postmortem blood specimens should be stored at low temperature to minimize the decrease in BAC during long periods of storage. But even with blood specimens from living subjects taken in sterile 10 ml Vacutainer tubes containing 1% NaF as preservative and kept at 4°C for several months, the concentration of alcohol decreased at a rate of $0.03\,g\,l^{-1}$ per month ($3\,mg\,100\,ml^{-1}$) on average. None of the bloods from living subjects showed increases in the concentration of alcohol and the rate of alcohol loss did not depend on the BAC present initially.

Postmortem Synthesis of Alcohol

Distinguishing antemortem ingestion of alcohol from postmortem synthesis has always been and still is a major dilemma for the forensic pathologist and toxicologist. Deciding whether alcohol was produced after death, e.g. by the action of microorganisms (bacteria or yeasts) on glucose or other substrates, is not easy. The opportunity for postmortem synthesis of alcohol exists from the moment of death to the time of the autopsy. After biological specimens are sampled, any further production of alcohol can be blocked by keeping the specimens cold (4°C) and treating them with ~2% sodium or potassium fluoride. However, to be effective as a preservative, the fluoride must be dissolved in the blood and this is not always easy to guarantee when specimens are clotted, e.g. intracranial hematoma.

Whenever gross decomposition and putrefaction of tissue has occurred, the question of postmortem synthesis of ethanol becomes particularly difficult. Factors to consider include the time since death to autopsy, indoor or outdoor conditions, bodies recovered from water or flames, temperature and humidity of ambient air, adequate ventilation, dehydration of the body etc. Many of these factors are listed in **Table 7**. Evidence of mummification, skin slippage, bloating, purging, discoloration, maggots, and bad-smelling corpses give strong indications of decomposition and putrefaction making it all the more likely that any alcohol present was produced after death. Addition of a preservative and enzyme inhibitor such as sodium or potassium fluoride after sampling will not help if alcohol has already been produced.

Bodies recovered from water and incinerated cadavers present special problems when the results

Table 7 Factors to consider when judging the significance of postmortem synthesis of alcohol

- Time between death and making the postmortem examination
- Condition and location of the body

    Indoors or outdoors
    Time of year
    Temperature and humidity of the environment
    Extent of traumatic injuries
    Abdominal trauma, e.g. ruptured stomach
    Incinerated body
    Body recovered from water
    Extent of decomposition/putrefaction

- Preservative (NaF) present in the blood samples at appropriate concentration (~2%)
- Number and kind of yeast or microorganisms present
- Availability of substrate for ethanol production: glucose, glycerol, lactate, ribose, various amino acids

of forensic alcohol analysis are interpreted. Both losses and increases in the concentration of alcohol are possible when a corpse is submerged in water for extended periods or burnt. Loss of alcohol from body fluids and tissue occur by dilution owing to high solubility in water as time after death increases. Environmental factors (temperature, salt or fresh water), long time lapses and extensive trauma to the body as well as lipolysis are other examples of complicating factors. An abundance of microorganisms present in the water might accelerate the production of ethanol after death.

The products of putrefaction besides ethanol include other alcohols, e.g. propan-1-ol, butan-1-ol, amyl alcohol, acetaldehyde, various fatty acids and their esters along with amines and other nitrogenous low-molecular-weight species including a number of gases, e.g. H_2S, NH_3, CH_4 and SO_2. Finding other peaks on the gas chromatogram besides ethanol and the internal standard confirms the presence of other volatile substances such as putrefaction products. Some investigators have suggested that identifying the presence of higher alcohols, e.g. propan-1-ol and butan-1-ol in the blood along with ethanol might be a useful indicator or marker for postmortem synthesis of ethanol. The concentrations of these other primary alcohols are usually much less than that of ethanol. Because propan-1-ol might be produced in decomposing tissue and typically in corpses retrieved from water, some toxicologists recommend that t-butanol is more appropriate as an internal standard for GC analysis. However, making duplicate determinations using both these internal standards and finding good agreement between the results speaks against the presence of propan-1-ol and thus any significant postmortem synthesis of ethanol. **Table 8** gives examples of various strategies available for judging whether alcohol determined in postmortem blood was produced after death by microbial activity.

A multitude of microorganisms are capable of producing alcohol from endogenous and exogenous substrates. The main candidate is glucose, which increases in concentration after death, especially in the liver and skeletal muscles. If necessary, the specimens of blood or tissue can be cultured and tests made for the presence of alcohol-producing species according to classical microbiological techniques. Low concentrations of ethanol ($<$20–30 mg 100 ml^{-1}) are more likely to be produced in postmortem blood than high alcohol concentrations ($>$200 mg 100 ml^{-1}).

Table 8 Methods used to judge whether postmortem synthesis of alcohol might have occurred

- Compare and contrast the concentrations of alcohol in blood taken from different parts of the vascular system (heart, femoral, cubital and jugular vein) look for wide discrepancies in results.
- Divide the blood sample into two portions one with preservative (2% NaF) and one portion without NaF and compare results of alcohol analysis.
- Make an initial analysis on blood without preservative then keep the specimen at room temperature in an airtight container for a few days before reanalysis. An appreciable increase or decrease in the concentration of alcohol suggests activity of microorganisms.
- Analyze and compare alcohol in femoral venous blood, bladder urine, vitreous humor and/or cerebrospinal fluid and look for abnormal concentration ratios.
- Culture the postmortem blood specimens and test the ability of any microorganisms present to produce ethanol by incubation with a suitable substrate.
- Identify the presence of other volatile substances in the blood specimen such as propan-1-ol, or butan-1-ol. These higher alcohols are produced in decomposing tissue and can therefore serve as indicators of postmortem synthesis of ethanol.
- Classify the extent of decomposition and putrefaction of the body from its appearance and distinctive color and smell.
- Determine the concentration of serotonin metabolites, 5-hydroxytryptophol to 5-hydroxyindoleacetic acid in urine. Ratios of 5HTOL/5HIAA above 15 indicate ingestion of alcohol some time before death.
- Consider the drinking habits of the deceased person and trace events 24 hours prior to death – was there any evidence of alcohol consumption?

Postmortem Diffusion

Postmortem diffusion relates to the movement or redistribution of alcohol and/or other drugs or their metabolites from one part of the body to another after death. The stomach, portal vein and liver are the main sites from which alcohol can diffuse into the pericardial and pleural fluid, and less readily into the chambers of the intact heart and extremities. These processes proceed for various periods of time after death. The more distant the tissue or fluid from the gut the lower the risk of diffusion artifacts occurring. Many factors influence postmortem diffusion including the physicochemical properties of the drug, tissue pH and the way the body was moved and handled after death. Whenever large quantities of alcohol remain unabsorbed in the stomach, postmortem diffusion needs to be considered when analytical results are interpreted. Furthermore, aspiration of vomit or

gastroesophageal reflux of stomach contents still containing a high concentration of alcohol will promote diffusion of alcohol into the upper airways and lungs.

Experiments in which high concentrations of alcohol (5% v/v or 40% v/v) were instilled into the stomachs of cadavers showed conclusively that alcohol spreads to many other sampling sites. However, the results from carefully sampling blood from the femoral vein were not compromised. The number of people who die with high concentrations of alcohol in their stomachs has not been well established although it is generally thought to be rather few. Finding a stomach–fluid alcohol concentration over 500 mg dl^{-1} ($5\,g\,l^{-1}$) and a considerably lower BAC, or UAC gives a strong indication of recent intake of alcohol before death.

Markers of Postmortem Synthesis of Alcohol

Methods are needed to decide whether a positive blood-alcohol concentration at autopsy actually reflects antemortem ingestion or microbial synthesis after death. Table 8 lists some of the current approaches to this problem although none are perfect and several should perhaps be used in combination.

A fairly new approach to this problem takes advantage of the metabolic interaction between ethanol and serotonin. When ethanol is being oxidized in the body, the coenzyme NAD^+ is converted into its reduced form NADH and the $NADH/NAD^+$ ratio in blood and tissue increases appreciably. This altered redox state in the liver disrupts many of the normal metabolic processes, including the metabolism of serotonin. The enzyme aldehyde dehydrogenase (ALDH), which is needed for the oxidation of acetaldehyde derived from ethanol, is also diverted from normal enzymatic pathways.

The biogenic amine serotonin (5-hydroxytryptamine) is deaminated by monoamine oxidase to produce an intermediate aldehyde, which is normally oxidized by ALDH to give an acid metabolite 5-hydroxyindoleacetic acid (5HIAA), which is excreted in urine. At the same time, a small fraction of the biogenic aldehyde ($\sim$1%) is reduced to an alcohol metabolite 5-hydroxytryptophol (5HTOL) by the action of ADH or aldehyde reductase. Studies have shown that the oxidative pathway, which normally dominates, switches to the reductive pathway during the metabolism of ethanol in part because of the availability of excess NADH and also the fact that ALDH is engaged in the oxidation of acetaldehyde into acetate. What all this means is that the ratio of 5HTOL/HIAA in urine, which is normally <15 pmol $nmol^{-1}$, increases in a dose-dependent manner during the time that alcohol is being metabolized. In postmortem toxicology, finding an elevated blood-alcohol concentration and a urinary 5HTOL/5HIAA ratio of less than 15 pmol $nmol^{-1}$ suggests that postmortem synthesis of ethanol has occurred in the blood samples. An elevated 5HTOL/5HIAA ratio in urine indicates that the person had been metabolizing alcohol before death.

Concluding Remarks

Many factors need to be considered when the results of analyzing alcohol in postmortem materials are interpreted for medicolegal purposes. Relating the measured postmortem BAC with an equivalent antemortem BAC and making a statement about whether a person had ingested alcohol before death is fraught with pitfalls. Translating the BAC into the amount of alcohol consumed is also subject to considerable uncertainty. The condition of the cadaver and especially any advanced state of decomposition/putrefaction may make the analysis and interpretation of BAC pointless. Under these conditions some or all of the alcohol might have been produced after death. Finding a postmortem BAC of 200 mg 100 ml^{-1} or more suggests that some alcohol at least was ingested prior to death even if the body was appreciably decomposed.

If emergency medical treatment was administered prior to death, e.g. infusion of solutions such as mannitol which causes an osmotic diuresis and helps to reduce swelling in the brain, or if intravenous fluids were given to counteract shock, this should be documented and considered when the results of alcohol analyses are interpreted. Such intensive care treatments can alter the concentrations of alcohol in the various body compartments and complicate interpretation of toxicological results. Moreover, mannitol is a sugar-alcohol and a good substrate for microbial synthesis of ethanol. The time delay between death and autopsy, and from autopsy to performing the laboratory analysis and the condition of the specimens on arrival (color, smell, viscosity, presence of clots) are also important to consider. Care is needed to ensure that the biological specimens were not contaminated by extraneous solvents at the place where the person died or when any emergency medical treatment was administered before death or at autopsy or in the laboratory where the specimens were analyzed.

Although blood from a femoral vein has become the preferred specimen for toxicological analysis of alcohol and other drugs, this is not always possible owing to severe injuries to the body. Scooping a specimen of 'blood' from the liquid pooled in the chest cavity is not good practice owing to the risk for internal or external contamination. These questionable procedures might falsify the results of alcohol analysis. Taking multiple specimens including blood, urine, vitreous humor or CSF provides additional information and gives a clearer picture about the person's BAC at the time of death. The concentrations of alcohol measured in urine, VH or CSF should not be routinely translated into the coexisting BAC owing to the wide individual variations in body fluid/blood ratios of alcohol. However, population average conversion factors are available and these might be considered if and when a conservative estimate of the coexisting BAC is required by analyzing one of the other body fluids or tissue in Table 3.

Much thought is needed when the results of postmortem blood-alcohol analysis are reported to the police because of the stigma attached to any suggestion that a deceased person was under the influence of alcohol. This becomes particularly important whenever fatal accidents on the roads or in the workplace are investigated and responsibility for the accident and insurance claims are made. Much debate and litigation erupts when claims of negligence are brought especially in mass transportation fatalities, which underscores the need to corroborate an elevated postmortem BAC by analyzing alternative biological specimens such as urine or vitreous humor or both. Finding a low blood-alcohol concentration $<0.5\,g\,l^{-1}$ ($<50\,mg\ 100\ ml^{-1}$) without any supporting evidence (e.g. positive urine or VH) or ancillary information concerning the person's drinking habits does not prove conclusively that the person had been drinking alcohol before death. The case history and the circumstances of the death including tracing important events by interviewing friends or relatives of the victim or by reading police records can sometimes provide useful hints about what to expect from the toxicological alcohol analysis.

In cases of extreme trauma when fragmented bodies are recovered (e.g. as might occur in aircraft accidents), when putrefaction is advanced, and when the body is burnt or recovered from water, an unequivocal answer about antemortem ingestion of alcohol from the analysis of postmortem blood specimens or tissue is probably not realistic. To paraphrase Professor Derrick J. Pounder '*Dead sober or dead drunk? Maybe hard to determine*.'

See also: **Alcohol**: Blood; Body Fluids; Breath; Interpretation; Congener Analysis. **Causes of Death**: Traffic Deaths. **Postmortem Examination**: Procedures and Standards. **Analytical Techniques**: Gas Chromatography.

Further Reading

Corry JEL (1978) Possible sources of ethanol ante- and post-mortem: its relationship to the biochemistry and microbiology of decomposition. *Journal of Applied Bacteriology* **44**:1–56.

Fleming R and Stotz E (1935) Experimental studies in alcoholism. *Archives of Neurology & Psychiatry* **33**:492–505.

Forrest ARW (1993) Obtaining samples at postmortem examination for toxicological and biochemical analyses. *Journal of Clinical Pathology* **46**:292–296.

Garriott JC (1998) Analysis of alcohol in postmortem specimens. In: Garriott JC (ed.) *Medicolegal Aspects of Alcohol*, 3rd edn, pp. 151–169. Tucson: Lawyers & Judges Publishing Co.

Harper DR and Corry JEL (1993) Collection and storage of specimens for alcohol analysis. In: Garriott JC (ed.) *Medicolegal Aspects of Alcohol Determination in Biological Specimens*, 2nd edn, pp. 145–169. Tucson: Lawyers & Judges Publishing Co.

Helander A, Beck O and Jones AW (1995) Distinguishing ingested ethanol from microbial formation by analysis of urinary 5-hydroxytryptophol and 5-hydroxyindoleacetic acid. *Journal of Forensic Sciences* **40**:95–98.

Jones AW and Pounder DJ (1998) Measuring blood-alcohol concentration for clinical and forensic purposes. In: Karch SB (ed.) *Drug Abuse Handbook*, pp. 327–356. Boca Raton: CRC Press.

Kalant H (1968) Interpretation of post-mortem ethanol concentrations. *Aerospace Medicine* **39**:633–667.

Levine BS (1999) *Principles of Forensic Toxicology*. Washington, DC: American Association of Clinical Chemistry.

Levine BS, Smith ML and Froede RC (1990) Postmortem forensic toxicology. *Clinics in Laboratory Medicine* **10**:571–589.

O'Neal CL and Poklis A (1996) Postmortem production of ethanol and factors that influence interpretation. *American Journal of Forensic Medicine and Pathology* **17**:8–20.

Pounder DJ and Jones AW (1998) Measuring alcohol postmortem. In: Karch SB (ed.) *Drug Abuse Handbook*. Boca Raton: CRC Press.

Pounder DJ and Smith DEW (1995) Postmortem diffusion of alcohol from the stomach. *American Journal of Forensic Medicine and Pathology* **16**:89–96.

Tagliaro F, Lubli G, Ghielmi S, Franchi D and Marigo M (1992) Chromatographic methods for blood alcohol determination. *Journal of Chromatography Biomedical Applications* **580**:161–190.

ANALYTICAL TECHNIQUES

Contents

Capillary Electrophoresis in Forensic Biology
Capillary Electrophoresis in Forensic Science
Gas Chromatography
Hyphenated Chromatographic-Spectroscopic Techniques
Mass Spectrometry
Microscopy
Presumptive Chemical Tests
Separation Techniques
Spectroscopic Techniques
Spectroscopy: Basic Principles

Capillary Electrophoresis in Forensic Biology

B R McCord, Clippinger Laboratories, Ohio University, Athens, OH, USA
E Buel, State of Vermont Forensic Laboratory, Waterbury, VT, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0772

Introduction

The development of methods for the amplification and detection of DNA fragments using the polymerase chain reaction (PCR) has resulted in rapid and dramatic advances in forensic DNA typing. Using the PCR it is possible to easily produce analytically significant amounts of a specified DNA product from trace quantities of DNA. In its forensic application, the PCR is used to demarcate and amplify known polymorphic sites on a distinct chromosome and produce discrete and easily characterized fragments of DNA.

The introduction of PCR-based forensic assays has also resulted in a need for efficient and automated procedures for analysis of the reaction products. This requirement has been the driving force behind the development of capillary electrophoresis (CE) methods for DNA analysis. In CE, DNA separations are performed in a thin 50–75 μm fused silica capillary filled with a sieving buffer. These capillaries have excellent capabilities to dissipate heat, permitting high electric field strengths to be used. As a result separations in capillaries are rapid and efficient. Additionally, the capillary can be easily manipulated for efficient and automated injections. Detection occurs via adsorption or fluorescence through a window etched in the capillary.

When used in short tandem repeat (STR) analysis, CE systems require specialized techniques. The high ionic strength of the PCR reaction mixture inhibits CE injection methods, and separations must be performed using highly viscous buffers. Detection of STRs is generally carried out by laser-induced fluorescence as adsorption techniques have poor sensitivity. Lastly, the serial nature of CE separations requires internal and external standardization to achieve highly precise measurements.

Theory of CE Separations

DNA fragments can be difficult to separate under normal CE conditions due to their virtually constant charge to mass ratio. Therefore, analyses are performed using a replaceable sieving matrix consisting of a water-soluble polymer dissolved in a suitable buffer. Such solutions are referred to as ‘physical gels’ as they are not chemically crosslinked. This fact makes them different from the crosslinked or ‘chemical gels’ used in slab gel analysis. The advantage of physical gels is that fresh gel solution can be pumped into the capillary at the conclusion of each analysis, thus limiting problems with carryover. Experiments carried out using a variety of physical gels have shown that with careful optimization of molecular weight and concentration, high resolution DNA separations can be produced.

Several different mechanisms have been postulated

to describe the separation of DNA in physical gels. These include transient entanglement coupling, Ogston sieving, and reptation. At low concentrations of polymer, separation takes place by means of a frictional interaction between the DNA and the polymer strands. This mechanism is known as transient entanglement coupling. At higher concentrations of polymer, individual polymer molecule strands begin to interact, producing a mesh. The polymer concentration at which this occurs is known as the entanglement threshold. Above the entanglement threshold, DNA fragments separate by sieving through transient pores created in the polymer mesh (**Fig. 1**). Fragments which are larger than the average pore size reptate or move in a snakelike manner through the mesh. The key to producing an acceptable separation is to specify a polymer concentration at which the size of these virtual pores approximates the radius of gyration of the DNA in solution.

To keep solution viscosity manageable, the polymer length must be kept to a minimum. Other characteristics of importance include the relative stiffness and polydispersity of the polymer. The key to producing an acceptable separation is to optimize the polymer molecular weight and concentration in the buffer solution using the resolution and mobility of the DNA as a guide to system performance. Other parameters such as column temperature, and applied electrophoresis voltage must also be tested. These variables affect resolution by affecting the rate of diffusion of the DNA as it moves down the capillary.

In uncoated capillary columns, charged silanol groups on the silica surface induce buffer ions to form a double layer along the capillary walls. Application of the electric field induces a flow of the bulk solution toward the negative electrode. This effect is known as electroosmotic flow (EOF). The magnitude and direction of the EOF is dependent on the number and type of active sites on the capillary surface and the pH of buffer. Electroosmotic flow can be considered a detriment to stable DNA separations because its velocity can change with each run, making peak migration times vary.

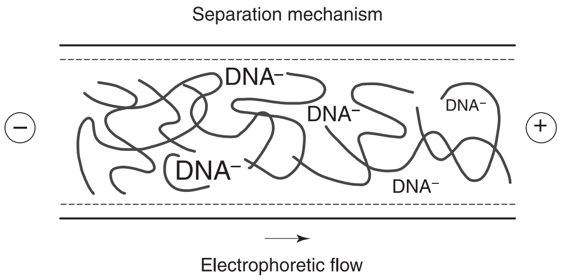


Figure 1 DNA is sieved through transient pores created in the polymer mesh. Smaller fragments are less impeded by the mesh and elute first.

To control EOF, capillary columns must be coated to mask these charged sites. Coatings can be classified into two types: dynamic coatings that must be periodically replenished or static coatings that are bonded to the capillary walls. Dynamic coatings are compounds added to the buffer, which mask active sites on the capillary walls. Many of the polymers developed for DNA separations perform this function, dynamically coating the capillary as well as sieving the DNA. Capillaries may also be washed with dilute HCl just prior to filling with buffer, to neutralize active sites on the capillary walls.

Static coatings are inert substances that are covalently bound to the internal walls of the capillary. Such coatings must be stable at the pH of analysis and free from contamination. The key factor in selecting a coating is its stability and lifetime under the conditions at which it is used. With periodic rinsing coated capillaries can last for months before failure.

Injection and Sample Preparation

There are two modes of injection in CE systems: hydrodynamic and electrokinetic. Hydrodynamic injections are performed using pressure to drive the samples into the capillary orifice. The solution viscosity, the capillary radius, and the applied pressure determine the volume of sample injected by this technique:

$$V = Pr^4 \pi t / 8 \eta L$$

where P is the pressure, r is the radius of the capillary, t is the time and η is the viscosity. Hydrodynamic injections are particularly well suited for quantitative analyses of DNA. When properly initiated, highly reproducible quantities of sample may be introduced onto the capillary. Standard deviations of DNA peak area using this technique have been shown to be 3%. However, the wide injection bands produced tend to limit resolution.

Electrokinetic injections are performed using an applied voltage to induce the sample to migrate into the capillary orifice. The quantity of material injected Q_{DNA} onto the capillary by this technique can be described by the following

$$Q_{DNA} = E\pi r^2 [DNA](\mu_{ep} + \mu_{eof})$$

where E is the field strength, r is the capillary radius, μ_{ep} is the electrophoretic flow and μ_{eof} is the electroosmotic flow. However, this equation must be modified as other ions present in solution will compete with DNA for the role as electrophoretic charge carriers. Thus, the quantity of DNA injected is also

a function of the ionic strength of the solution, and the total quantity injected:

$$T_{\mathrm{DNA}} = Q_{\mathrm{DNA}} / \sum Q_i$$

Electrokinetic injections produce narrow injection zones but are highly sensitive to the sample matrix. For example, the electrokinetic injection of PCR products into a capillary is inhibited by the high salt content of the sample matrix (>50 mM Cl^-). These small ions also tend to be selectively injected into the capillary. To overcome this problem, PCR samples can be purified by means of dialysis, spin columns or ethanol precipitation. The dialysis step appears to be the most effective for removing excess salt, whereas the spin columns are more effective at removing primer peaks, enzyme and dNTPs. Dilution of the sample in water or deionized formamide is another technique for reducing ionic strength. The DNA signal can then be selectively enhanced utilizing fluorescence detection.

The above steps increase quantity of sample injected by removing interferences or by sharpening the injection zone in a process known as stacking. Stacking, also called field amplified injection, occurs when the ionic strength of the sample zone is lower than that of the buffer. As the current through the system is constant, the lack of charge carriers in the sample zone produces a strong electric field that ends abruptly at the interface between the sample zone and the buffer inside the capillary (**Fig. 2**). DNA molecules mobilized by this field move rapidly towards the capillary as the injection voltage is applied and 'stack' in a narrow zone at the interface. Stacking allows a large sample zone to be loaded onto the capillary with a minimum of band broadening. Stacking also aids in producing efficient separations. With sharp injection zones, shorter capillaries and less gel media is required to effect a separation.

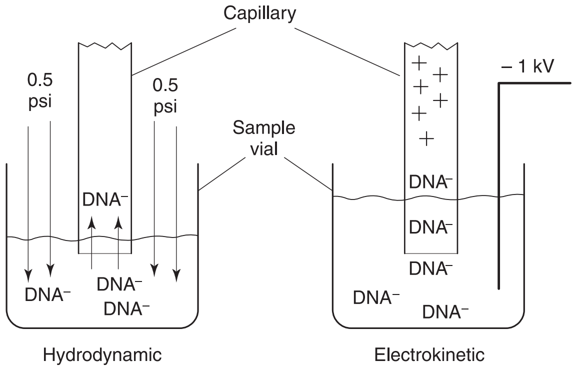


Figure 2 The two modes for injection for CE. In hydrodynamic injections, pressure is used to force sample into the capillary. In electrokinetic injections, an applied voltage causes the DNA to migrate into the capillary. Figure courtesy of Dr Butler, GeneTrace Systems.

Detection and Data Analysis

Early capillary electrophoretic separations of PCR products utilized UV absorbance detection, and required extensive purification and concentration of the samples prior to analysis. The relatively short path length and the dispersion produced by the capillary walls limited sensitivity. Laser-induced fluorescence (LIF) solved this problem by focusing an intense beam of light directly onto the sample volume within the capillary. Detection enhancements of 400-fold or more have been achieved using LIF as compared to UV detection.

Fluorescence detection of DNA is achieved by derivatizing the DNA using dyes to produce fluorescent adducts. Native DNA fragments can be detected using fluorescent intercalating dyes. These dyes bind to the DNA molecule by binding to the DNA helix, affecting the configuration of the aromatic rings of the dye and producing a large enhancement of the fluorescence signal. Additionally, intercalating dyes help to minimize effects of DNA structure on migration rate, resulting in better estimates of fragment lengths. The low background fluorescence of the uncomplexed dyes allows them to be added directly to the CE buffer.

Dye molecules may also be covalently bound to the DNA fragments. This procedure is commonly used to label single-stranded DNA by attaching a dye molecule to the 5′ end of each primer. Typically, only one primer is labeled with a fluorescent dye to avoid doublet bands. Upon completion of the PCR, all of the target DNA molecules are labeled with a fluorophore. By using a variety of different dyes in a single multiplexed reaction, different loci can be targeted, amplified, and labeled with specific dyes. The dyes used in these reactions absorb at similar wavelengths, but emit at different wavelengths. Thus a single laser can be used to excite four or more dyes. A multichannel analyzer is then used to identify the specific PCR product by means of the wavelength of emission of the bound dye.

The development of methods for data analysis by CE is of particular importance in the examination of PCR products. Precise and reliable methods must be developed for product analysis. Slab gel methods permit the analysis of multiple samples run concurrently. CE is a serial technique; samples are run one at a time. Thus, comparison of multiple samples requires the addition of internal standards to correct for the inevitable variations in injection, temperature, and current. This observation is particularly relevant in

quantitative methods where variations in sample injection can limit the usefulness of the technique.

Size estimates of PCR products can be performed by interpolation of product size based on the migration of one or more internal standards. For products in the size range 100–400 bp, a linear relationship exists between size and migration time. The size of larger products may also be estimated using nonlinear curve-fitting algorithms. Specialized instrumentation has been developed specifically for DNA analysis which utilizes internal standards that have been labeled with a different fluorescence dye than that of the product. For such systems specific algorithms have been developed to deconvolute the fluorescence signals and perform size estimates

Short Tandem Repeats

Short tandem repeats or STRs are tandemly repeated nucleotide sequences 2–6 base pairs in length. The number of repeated sequences varies between individuals and results in a high degree of length polymorphism. STRs are abundant throughout the human genome, occurring at an average rate of every 6–10 kilobases. The most common form of STR is the dinucleotide repeat. These loci are unsuitable for most forensic analyses due to the presence of PCR artifacts known as 'stutter' bands. This phenomenon is manifested by the appearance of extra peaks one or more repeat units away from the main product and is presumably caused by enzyme slippage during the amplification process. Tetrameric repeats tend to produce less stutter than the di- or trimeric repeats and much work has been done to validate their use in forensic casework. A core set of 13 loci has been established by the Federal Bureau of Investigation for use in the Combined DNA Index System (CODIS) (**Table 1**).

Table 1 13 STR loci approved for use with CODIS[a]

STR locus	*Chromosome*	*Repeat motif*	*Number of repeats*[b]
FGA	4	CTTT[c]	18–30
VWA	12	TCTA	11–22
D3S1358	3	TCTA	11–20
D21S11	21	TCTA	25–36
D8S1179	8	TATC	8–19
D7S820	7	GATA	6–15
D13S317	13	TATC	8–15
D5S818	5	AGAT	7–16
D16S539	16	GATA	5–15
CSF1P0	5	AGAT	6–15
TPOX	2	AATG	6–13
THO1	11	TCAT	3–11
Amelogenin[d]	X,Y		

[a] Data obtained from STRbase, published by NIST, http://ibm4.carb.nist.gov:8800/dna/home.htm and from the Profiler and Profiler+ users manuals, Perkin-Elmer, Foster City, CA.
[b] The range of repeats is approximate as new alleles are constantly being discovered.
[c] FGA as well as other loci in this list have complex patterns of repeats. The most common is given.
[d] Amelogenin is a sex-linked marker which contains a six base deletion in the X chromosome.

Sample Preparation

Purified template DNA can be amplified to yield products from a single STR locus or multiple primers can be added to a single reaction mixture to products from multiple STR loci. Cocktails are commercially available for multiplex PCR reactions that include primers for as many as 15 different STR loci. The products of these reactions are labeled with as many as three different fluorescent dyes. Both the number of dyes and coamplified STRs are certain to increase. The amplified products can be denatured in de-ionized formamide that is commercially obtained or prepared on site by using an ion-exchange resin. Care must be taken when using these resins as ionic species in the formamide can inhibit the injection process. For this reason, the conductivity of the formamide should be tested and control samples should be run prior to routine analysis. The quality of the formamide used in sample preparation can also affect column longevity.

To yield denatured fragments ready for electrophoresis, amplified product is added to de-ionized formamide, heated at 95°C and then snap cooled in an ice bath. Some investigators have substituted de-ionized water for formamide and have achieved results similar to those obtained using formamide, however long-term DNA stability is compromised when using water as a diluent. As mentioned earlier, filtration or dialysis can be performed on the PCR product prior to mixing with formamide to increase the amount of sample injected, however, these procedures may be unnecessary for routine analysis.

Analytical Separation

Electrokinetic injection is routinely employed to apply samples onto the capillary column. Injection time may be varied within certain ranges to affect the amount of sample applied to the column without adversely affecting resolution. Varying the injection time from 1 to 10 s has been shown to increase sample input while maintaining the resolution of the system. A particular advantage of CE is the ability to quickly reanalyze overly dilute samples by simply increasing the injection time. Samples that contain excess amplified product can be either reinjected for less time or simply diluted in a mixture of formamide and size standard and rerun.

The separation media employed for analysis are typically polydimethyl acrylamide, hydroxyethyl cellulose, linear polyacrylamide or commercially prepared polymer solutions. Most forensic laboratories have opted to purchase prepared polymer solutions for reasons of quality control and simplicity of use. Varying the polymer concentration, through the purchase of the appropriate separation media or by preparation, allows the user to fine-tune resolution.

The STRs under current study contain alleles that generally differ by a four-base repeat unit, however variants which contain deletions of one or two bases can occur in these systems. As a result it is important to design separation systems which can also resolve variant alleles. In situations where increased resolution is necessary, column length or polymer concentration can be increased, however both of these procedures can greatly increase migration times, a concern for laboratories with large numbers of samples.

Monitoring the resolution of a system allows the investigator to recognize degradation in column performance or inappropriate sample preparation. As the column ages through continued use or inadequate maintenance, sample resolution will deteriorate. Samples prepared in formamide that has not been sufficiently de-ionized will also show poor resolution. Resolution can be calculated using the standard equation:

$$R = (t_2 - t_1)/\frac{1}{2}(w_1 + w_2)$$

where t is the migration time of the peak and w is the peak width. Since the peak at base line is difficult to determine, the shape of the CE peak can be assumed to be gaussian with a width of 4σ, and the above equation can be converted to:

$$R = [2(\ln 2)]^{1/2}(t_2 - t_1)/(wh_1 + wh_2)$$

where wh is the peak width at half height.

System resolution can also be quickly evaluated using a mixture containing STR alleles that vary by one base pair. The STR system THO 1 has a variant (allele 9.3) that is one base pair less than expected. This allele can be mixed in equal proportions with the nonvariant allele 10. After this mixture is run on the capillary, the evaluation of the relative height of the peaks versus the valley point between them yields a ratio that can be monitored to obtain the relative resolution of the system. Most forensic systems under current study do not provide baseline resolution of STRs that differ by one base pair, and hence monitoring the relative valley value can provide a handle on the relative performance of the system.

Genotyping

Multiple STR loci can be determined during a single analysis by adjusting the fragment length of each PCR product and by labeling sets of different primers with dyes that fluoresce at differing wavelengths. Primers are carefully designed to produce DNA with allele sizes that do not overlap. This permits the analysis of 3–4 STR loci which are labeled with the same color. **Figure 3** illustrates this procedure by showing an

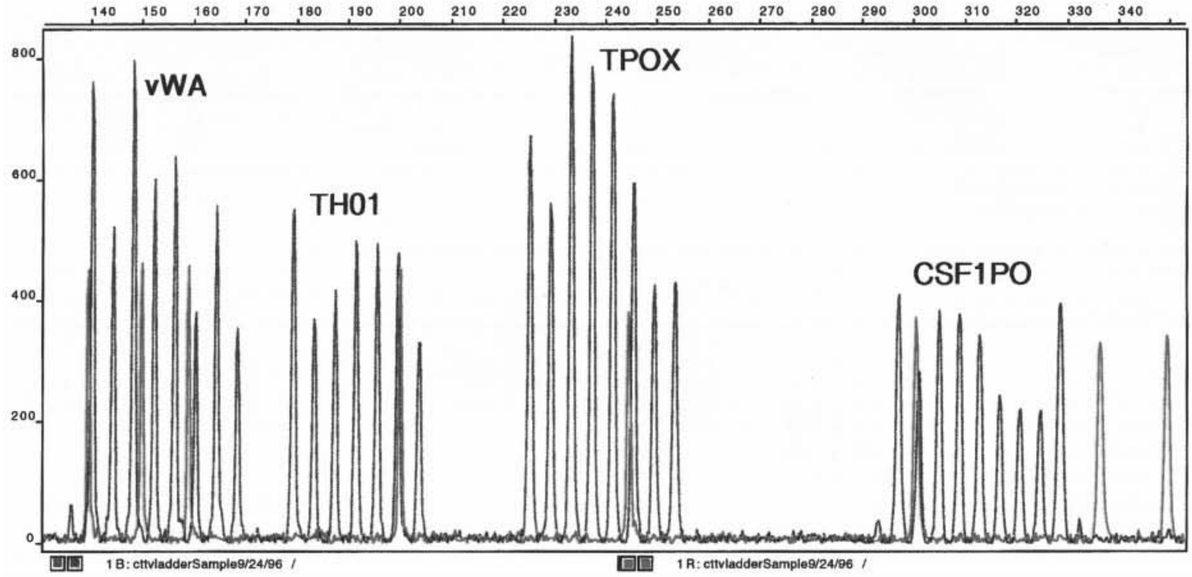


Figure 3 (see color plate 7) An allelic ladder from the Geneprint cttv multiplex (Promega) consisting of alleles from three genetic loci, CSF1P0, TPOX, THO1, and vWA. The internal standard (labeled in red) is a Genescan 350 ROX (Perkin-Elmer) size standard. Analyzed using an PE/ABI 310 capillary electrophoresis system with laser induced fluorescence detection and a hydroyxethyl cellulose sieving matrix.

allelic ladder consisting of all possible alleles from three different STR loci amplified together and labeled with a blue dye. Then, by using mixed sets of colored primers, 10 or more loci may be analyzed simultaneously (**Fig. 4**). This task is performed using CE instruments equipped with detector arrays capable of analyzing all dyes simultaneously.

The peaks resulting from this analysis can be genotyped through the use of software supplied by the instrument manufacturer. Typically, electrophoresis is conducted by combining amplified products with an internal size standard that is derivatized with a fluorescent dye that is different from those used to label the STR loci. Figure 3 illustrates this technique. In this figure the red size standard provides an internal reference to standardize the electrophoretic run and permits the calculation of the base pair sizes of the detected peaks. The calculated sizes can be compared to those sizes obtained from the allelic ladders, which are run separately. The allelic ladder is used to calibrate a series of electrophoretic runs, much like the internal standard used to standardize a single run.

The resulting data are compared to tables containing the frequency of each allele in a target population. The frequency of a particular profile can then be calculated by multiplying together the component

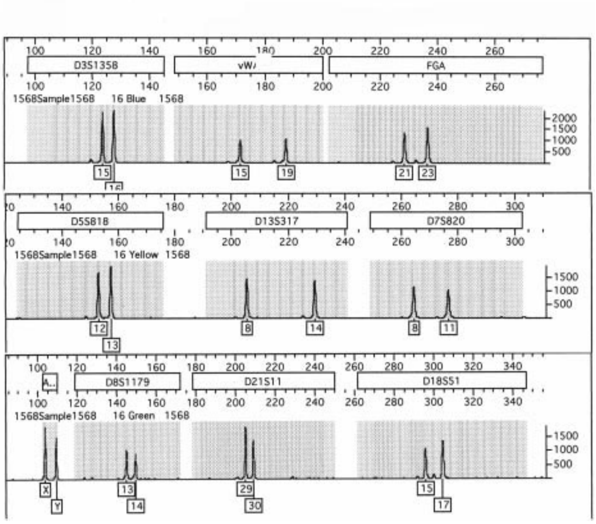


Figure 4 A single analysis of a blood sample amplified using the Profiler+® (Perkin-Elemer) PCR amplification kit. The results have been split into three panels to aid in analysis. Each gray panel indicates a different STR locus and is identified above the panel. The dark gray zones within the lighter gray indicate potential locations of alleles for each locus. Note the small stutter peaks to the left of certain of the larger peaks. Conditions: PE/ABI 310 genetic analyzer using POP4 sieving matrix and laser induced fluorescence detection.

frequencies calculated for each locus. The resultant frequencies can be quite small. For example, one commercial set of 10 loci has a probability of identity of 1×10^{-11}.

Mixture Analysis

Mixtures, samples that contain DNA from more than one individual, must be anticipated in the analysis of forensic specimens. These specimens may be a composite of body fluids from different individuals and will produce complex DNA profiles. To complicate the analysis of mixtures, STR patterns from one individual may contain imbalanced peaks and PCR artifacts known as stutter.

Stutter peaks are amplified products resulting from the 'slippage' of DNA polymerase during amplification where the enzyme and growing DNA chain are out of alignment with the target DNA. The resulting fragment is usually 4 bp less than the true allelic peak, although weaker signals consisting of sets of peaks four bases apart may also be seen. Some loci have yielded stutter peaks of $>10\%$ of the height of the true peak. In the interpretation of mixtures, the possibility of stutter peaks must be taken into account and interpretations should be adjusted based on the amount of stutter observed at a particular locus. Typically, a mixture is suspected when peak heights rise above typical stutter values for a particular locus.

Another problem in the interpretation of mixtures is that peaks obtained from a heterozygous locus may vary by as much as 30%. Deviations of this size, although uncommon, must be considered in the evaluation of mixtures. Differences in the expected peak ratio in a sample that presents a heterozygous pattern, can indicate a mixed sample whose alleles have co-electrophoresed. Fortunately, when multiple allelic systems are evaluated, other loci may show three or four peaks, establishing the specimen as a mixture, and can be used to determine if the altered ratio could be due to an overlapping allele. **Figure 5** illustrates the analysis of a sample of mixed DNA.

The balance observed between loci is an additional consideration in the assessment of mixtures. Although commercially available STR amplification kits attempt to achieve a balance, as the size of the template required for amplification increases, the amount of amplified product typically decreases. In situations where degraded samples are used, shorter PCR products will predominate as few of the longer fragments of template have survived. Under such circumstances, peaks that represent a minor component in a short locus may not be present in other

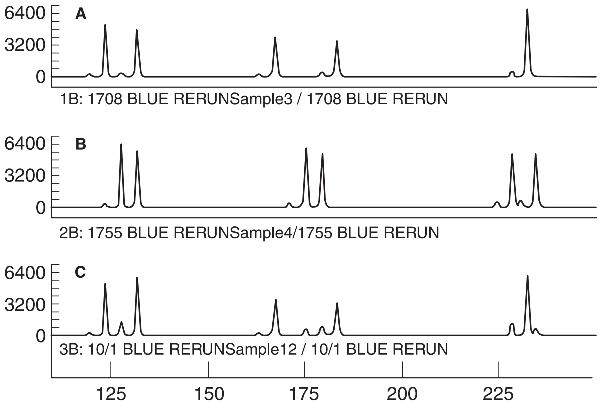


Figure 5 The analysis of a mixed sample of DNA using the AmpFlSTR blue STR kit (Perkin-Elemer). Panels A and B contain individual samples. Sample C is a mixture with sample B as the minor component. Note the effect of stutter peaks on the relative size of the alleles. Conditions as in Figure 4.

larger-sized loci. Nonallelic peaks occasionally also cause problems in the interpretation of a mixed sample. These peaks may be the result of free dye, unbound dye, electrical interferences or other sample artifacts. There are also PCR artifacts which occur resulting in a peak one base pair less than the true allelic peak. In most STR systems the amplification process promotes the nontemplate addition of a nucleotide (usually an A) to the end of the PCR product. This yields an amplicon that has been increased in size by one base pair. Under some conditions, such as excessive amount of template DNA or Taq inhibitors, the complete conversion of all the products to the n+1 state may not occur. Often this condition can be rectified by increasing the extension time for the amplification.

Finally, mutations or rare genetic events, may give rise to unusual profiles that can be mistaken as a mixture. In addition, mutations at primer sites may lead to genotyping variations at particular loci, due to differences in the location of primer annealing sites used by the various manufacturers of STR amplification kits.

Databasing

For many years most USA DNA laboratories did not pursue the analysis of cases suitable for DNA analysis unless standards for comparison were obtained. With the introduction of a national DNA database, this policy has changed. Now many laboratories are preparing to profile specimens in cases where no suspect exists in the hope of identifying the suspect through a database search. The FBI has established CODIS as a database to allow for the exchange and comparison of DNA profiles. All 50 states have legislation that enables them to collect DNA specimens from individuals convicted of certain crimes. When profiled, CODIS will act as the repository for the convicted offender profiles so that laboratories linked to the database can access the files. In addition to the database of convicted offenders, CODIS will also permit the comparison of profiles generated from crime scene samples that would allow investigators to link crimes to repeat offenders. Many crimes have been solved or linked through database searches, which in the USA have been primarily RFLP (restriction fragment length polymorphism) based. With the advent of STRs, and given the relative ease of analysis compared to RFLP, more samples of both convicted offenders and crime scenes will be profiled and entered into the database.

Analysis of Mitochondrial DNA

There are circumstances in forensic analysis in which there is insufficient nuclear DNA to perform PCR. These cases involve samples such as shed hairs or those that are highly degraded. In these circumstances there may still be enough mitochondrial DNA to permit PCR amplification. The DNA present in mitochondria is approximately 16 000 bases long and contains a section known as the control region that contains a number of polymorphic sites which are usually point mutations. In this procedure, certain hypervariable segments of the control region are PCR amplified and sequenced. These sequences are then compared to a known standard in order to identify polymorphic sites.

CE is used in this process to determine if the amplified product is present in sufficient quantity and purity for the sequencing step to be practical. A small portion of the amplified product is analyzed in its native state using a short 27 cm capillary at 15 000 V. In the analysis, an intercalating dye is added to the detector to provide a fluorescent product and a result is obtained in under 4 min. The peak intensity of the amplified DNA is compared to an internal standard to determine the quantity of amplified material. **Figure 6** illustrates this separation. The electropherogram is also checked to determine if any contaminants such as primers or extraneous amplified product is present in the sample. These materials can interfere with the sequencing reactions and reduce the quality of the result. The results of this analysis are used to adjust the amount of template used for the sequencing reaction. The products are then analyzed on gel-based sequencers. However, as sequencers based on CE become more widespread, both the product analysis and the sequence analysis will be performed via CE.

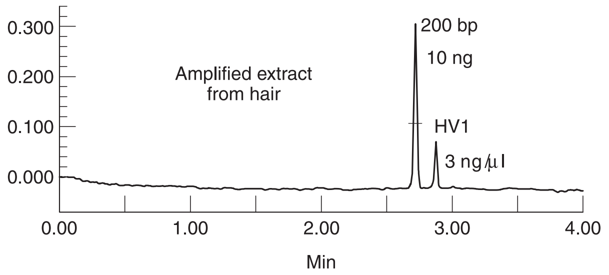


Figure 6 The analysis and quantitation of mitochondrial DNA in under four minutes. The quantity of DNA produced is determined with reference to the internal standard. Conditions: Beckman PACE capillary system with laser induced fluorescence. 27 cm DB-17 capillary run at 15 kV and filled with 1% hydroxyethyl cellulose and 50 ng/ml YO-PRO-1 in TBE buffer. Figure courtesy of Dr. John Butler, GeneTrace Systems.

There are also a number of other methods for detection of point mutations in mitochondrial DNA for which CE can be used. These include single-strand conformation polymorphism (SSCP) and PCR–RFLP. These techniques can provide a quick comparison of known and questioned samples of the amplified products. If the patterns produced do not match, then an exclusion has occurred and it is not necessary to perform the sequencing step.

Future Applications

Unlike slab gel-based systems, CE instruments must inject samples in a sequential fashion. However, it is possible to perform the capillary experiment using a bundle or array of individual capillaries to inject multiple numbers of samples at the same time. Such systems can inject 96 samples or more at the same time. Then, using software algorithms, the migration times from each individual capillary are scanned, synchronized using internal lane standards, and the data are reported. Capillary array electrophoretic systems can produce tremendous amounts of data in a relatively short time.

Recognizing the increased complexity required by the capillary array systems, a number of researchers are developing smaller, more compact systems based on microchip technology. By using photolithography, multiple channels can be etched into silicon wafers and the entire capillary array can be placed on a silicon chip. Tightly focused sample injections on a microchip permit fast electrophoretic separations using relatively short channels. A further advantage of this technique is that sample preparation and detection apparatus can be built into the chip design. Thus the entire process from DNA extraction to PCR to separation to detection can be integrated into a single device.

Conclusions

Capillary electrophoresis is a technique which provides the DNA analyst with much flexibility. CE systems utilize replaceable physical gels which are pumped into the capillary at the beginning of each analysis. DNA quantitation, genotyping and sequencing are all possible using this technique. Sample injection, separation and analysis can be easily automated. Multichannel fluorescence detection permits multiplex PCR reactions to be simultaneously analyzed, greatly conserving precious forensic samples. The resulting data can be analyzed to detect the presence of mixtures, processed, and stored in a database known as CODIS. Present systems under development will utilize arrays of capillaries to increase sample throughput. Future integration of the entire process of DNA analysis is possible using CE due to the distinct advantages of this technique.

See also: **Analytical Techniques**: Capillary Electrophoresis in Forensic Science. **Deoxyribonucleic Acid**: Basic Principles; Restriction Fragment Length Polymorphism; Polymerase Chain Reaction; Polymerase Chain Reaction-Short Tandem Repeats; Future Analytical Techniques; Significance; Databanks; Mitochondrial Deoxyribonucleic Acid.

Further Reading

Buel E, Schwartz MB and LaFountain MJ (1998) Capillary electrophoresis STR analysis: comparison to gel-based systems. *Journal of Forensic Sciences* **43**:164–170.

Edwards A, Civitello A, Hammond HA and Caskey CT (1991) DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *American Journal of Human Genetics* **49**:746–756.

Hammond HA, Jin L, Zhong Y, Caskey CT and Chakraborty R (1994) Evaluation of 13 short tandem repeat loci for use in personal identification applications. *American Journal of Human Genetics* **55**:175–189.

Hammond HA and Caskey CT (1994) Human DNA fingerprinting using short tandem repeat loci,. *Methods in Molecular and Cellular Biology* **5**:78–86.

Heller C (ed) (1997) *Analysis of Nucleic Acids by Capillary Electrophoresis*. Wiesbaden, Germany: Friedr. Vieweg.

Isenberg AR, Allen RO, Keys KM, Smerick JB, Budowle B and McCord BR (1998) Analysis of two multiplexed short tandem repeat systems using capillary electrophoresis with multiwavelength fluorescence detection. *Electrophoresis* **19**:94–100.

Lazaruk K, Walsh PS and Oaks F *et al.* (1998) Genotyping of forensic short tandem repeat (STR) systems based on sizing precision in a capillary electrophoresis instrument. *Electrophoresis* **19**:86–93.

Mansfield ES, Robertson JM, Vainer M *et al.*, (1998) Analysis of multiplexed short tandem repeat (STR)

systems using capillary array electrophoresis. *Electrophoresis* **19**:101–107.
National Research Council (1996) *The Evaluation of Forensic DNA Evidence*. Washington, DC: National Academy Press.
Salas-Solano O, Ruiz-Martinez MC, Carrilho E, Kotler L and Karger BL (1998) A sample purification method for rugged and high-performance DNA sequencing by capillary electrophoresis using replaceable polymer solutions. B. Quantitative determination of the role of sample matrix components on sequencing analysis. *Analytical Chemistry* **70**:1528–1535.
Swerdlow H, Jones BJ and Wittwer CT (1997) Fully automated DNA reaction and analysis in a fluidic capillary instrument. *Analytical Chemistry* **69**:848–855.
Ulfelder K and McCord BR (1996) Capillary Electrophoresis of DNA. In: Landers JP (ed.) *Handbook of Capillary Electrophoresis*, pp. 347–378. New York: CRC Press.
Wallin JM, Buoncristiani MR, Lazaruk K, Fildes N, Holt CL and Walsh PS (1998) TWGDAM validation of the AmpFlSTR BluePCR Amplification Kit for forensic casework analysis. *Journal of Forensic Science* **43**:854–870.
Walsh PS, Fildes NJ and Reynolds R (1996) Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA. *Nucleic Acids Research* **24**:2807–2812.
Woolley AT, Hadley D, Landre P, deMello AJ, Mathies RA and Northrup MA (1996) Functional integration of PCR amplification and capillary electrophoresis in a microfabricated DNA analysis device. *Analytical Chemistry* **68**:4081–4086.

Capillary Electrophoresis in Forensic Science

F Tagliaro and **VL Pascali**, Institute of Forensic Medicine, Catholic University of the Sacred Heart, Rome, Italy

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0420

Introduction and Background

Capillary electrophoresis (CE) (also known as high-performance capillary electrophoresis (HPCE)) was devised about 20 years ago and, essentially, can be regarded as the instrumental optimization in capillaries of traditional electrophoretic techniques. This somewhat elaborate process of 'evolution' took about a decade to accomplish and by the early 1990s the technique was virtually ready for significant practical applications. Since then, CE has had extensive coverage in a variety of scientific domains, including chemistry, biochemistry, pharmacology, toxicology, and biomedicine. A survey in December 1998 discovered more than 5000 internet web pages matching the key words capillary electrophoresis, an increase of 3000 since the previous year.

One of the reasons leading to this clear success is that CE is indeed a very versatile technique. It enables separation of a variety of molecules – from small inorganic ions to huge biopolymers – with much the same equipment, although using separation modes based onto completely different physicochemical mechanisms. Another is that highly sophisticated, user friendly, and very effective CE instruments have become commercially available at affordable costs.

Typically, CE separations are highly efficient (up to millions of theoretical plates), sensitive (in terms of mass and concentration), fast (usually <30 min), simple (no derivatization needed), and require extremely small amounts of solvents (few tens of ml per day of running buffer) and other consumables (e.g. capillaries).

Recently, there has been great interest in CE in the world of forensic science with the publication of review articles and journals' special issues on this subject.

The purpose of this article is to introduce some basic concepts of CE for the general forensic science audience and to provide a comprehensive overview on the most outstanding applications of this technique to a variety of forensic disciplines, includes forensic toxicology, forensic biology/biochemistry (DNA profiling, analysis of proteins and other biological compounds of forensic interest), explosive and gunshot residue analysis, and ink analysis.

Instrumentation

A CE instrument, also called a capillary electropherograph, is so simple, that 'home made' instrumentation can be developed for specialized applications, research, or tutorial purposes. However, the best analytical performances, in terms of reproducibility, accuracy and automation can only be achieved with commercial instrumentation.

Briefly, a capillary electropherograph (**Fig. 1**) consists of an injection system, a capillary, a high voltage source (10–30 kV), electrodes and electrode jars, and a detector.

In CE, the separation occurs in a capillary and this is also the compartment where injection and detection takes place. Ideally a capillary should be chemically and physically resistant, precisely machined with

systems using capillary array electrophoresis. *Electrophoresis* **19**:101–107.
National Research Council (1996) *The Evaluation of Forensic DNA Evidence*. Washington, DC: National Academy Press.
Salas-Solano O, Ruiz-Martinez MC, Carrilho E, Kotler L and Karger BL (1998) A sample purification method for rugged and high-performance DNA sequencing by capillary electrophoresis using replaceable polymer solutions. B. Quantitative determination of the role of sample matrix components on sequencing analysis. *Analytical Chemistry* **70**:1528–1535.
Swerdlow H, Jones BJ and Wittwer CT (1997) Fully automated DNA reaction and analysis in a fluidic capillary instrument. *Analytical Chemistry* **69**:848–855.
Ulfelder K and McCord BR (1996) Capillary Electrophoresis of DNA. In: Landers JP (ed.) *Handbook of Capillary Electrophoresis*, pp. 347–378. New York: CRC Press.
Wallin JM, Buoncristiani MR, Lazaruk K, Fildes N, Holt CL and Walsh PS (1998) TWGDAM validation of the AmpFlSTR BluePCR Amplification Kit for forensic casework analysis. *Journal of Forensic Science* **43**:854–870.
Walsh PS, Fildes NJ and Reynolds R (1996) Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA. *Nucleic Acids Research* **24**:2807–2812.
Woolley AT, Hadley D, Landre P, deMello AJ, Mathies RA and Northrup MA (1996) Functional integration of PCR amplification and capillary electrophoresis in a microfabricated DNA analysis device. *Analytical Chemistry* **68**:4081–4086.

Capillary Electrophoresis in Forensic Science

F Tagliaro and **VL Pascali**, Institute of Forensic Medicine, Catholic University of the Sacred Heart, Rome, Italy

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0420

Introduction and Background

Capillary electrophoresis (CE) (also known as high-performance capillary electrophoresis (HPCE)) was devised about 20 years ago and, essentially, can be regarded as the instrumental optimization in capillaries of traditional electrophoretic techniques. This somewhat elaborate process of 'evolution' took about a decade to accomplish and by the early 1990s the technique was virtually ready for significant practical applications. Since then, CE has had extensive coverage in a variety of scientific domains, including chemistry, biochemistry, pharmacology, toxicology, and biomedicine. A survey in December 1998 discovered more than 5000 internet web pages matching the key words capillary electrophoresis, an increase of 3000 since the previous year.

One of the reasons leading to this clear success is that CE is indeed a very versatile technique. It enables separation of a variety of molecules – from small inorganic ions to huge biopolymers – with much the same equipment, although using separation modes based onto completely different physicochemical mechanisms. Another is that highly sophisticated, user friendly, and very effective CE instruments have become commercially available at affordable costs.

Typically, CE separations are highly efficient (up to millions of theoretical plates), sensitive (in terms of mass and concentration), fast (usually <30 min), simple (no derivatization needed), and require extremely small amounts of solvents (few tens of ml per day of running buffer) and other consumables (e.g. capillaries).

Recently, there has been great interest in CE in the world of forensic science with the publication of review articles and journals' special issues on this subject.

The purpose of this article is to introduce some basic concepts of CE for the general forensic science audience and to provide a comprehensive overview on the most outstanding applications of this technique to a variety of forensic disciplines, includes forensic toxicology, forensic biology/biochemistry (DNA profiling, analysis of proteins and other biological compounds of forensic interest), explosive and gunshot residue analysis, and ink analysis.

Instrumentation

A CE instrument, also called a capillary electropherograph, is so simple, that 'home made' instrumentation can be developed for specialized applications, research, or tutorial purposes. However, the best analytical performances, in terms of reproducibility, accuracy and automation can only be achieved with commercial instrumentation.

Briefly, a capillary electropherograph (**Fig. 1**) consists of an injection system, a capillary, a high voltage source (10–30 kV), electrodes and electrode jars, and a detector.

In CE, the separation occurs in a capillary and this is also the compartment where injection and detection takes place. Ideally a capillary should be chemically and physically resistant, precisely machined with

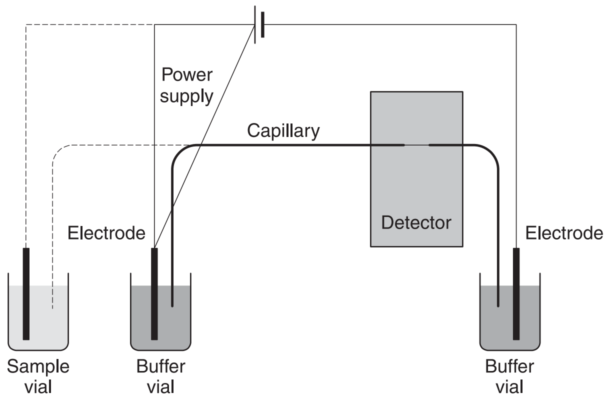


Figure 1 Schematic representation of a capillary electropherograph; interrupted lines indicate the different capillary and electrode locations during electrokinetic injection.

narrow internal diameters (typically i.d. 20–100 μm; length of 20–100 cm), not prone to adsorb solutes, transparent to UV–visible radiation, and should have a high thermal conductivity and low cost. Fused silica capillaries with an external protective polyimide coating (similar to those used in capillary gas chromatography) meet almost all these requirements and are consequently a standard choice in CE. Capillaries can be internally 'uncoated', having the inner silica surface in direct contact with buffers and solutes, but, under some experimental conditions, these may adsorb analytes. Alternatively, the capillaries can be internally coated with a thin layer of polymers (adsorbed or chemically bound) shielding the silica surface from interactions with solutes. Capillaries can be also filled with gels, mimicking slab gel electrophoresis, or with particles carrying a stationary phase, reproducing a chromatographic system.

At the capillary wall, a crucial phenomenon, such as the so-called electroosmotic flow (EOF), takes place. It is generated because the inner wall of the fused silica capillary exposes a great number of silanol groups to the buffer, which are negatively ionized (as SiO^-) at pH values higher than about 4 and which interact electrostatically with the ions in solution. Thus, cations, in the case of silica capillaries, are attracted and concentrated at the interface with the capillary wall and, when a potential difference is established between the ends of the capillary, migrate towards the cathode (negative electrode) drawing water with them by osmosis. This generates a flow of liquid inside the capillary which is named EOF. A peculiarity of this liquid flow is that, being generated at the capillary wall, it has a flat profile (piston-like). This limits the band-broadening that occurs during CE separations, which is a typical drawback of capillary liquid chromatography, where the flow of the mobile phase is generated by pressure. Because of its nature, EOF is highly affected by pH, ionic strength and composition of the buffer. To abolish (or reverse) the EOF the capillary wall can be coated with 'modifiers', which can be physically adsorbed, chemically bound or simply added to the running buffer (dynamic coating). The most common coatings are polyacrylamide, cellulose, polyvinyl alcohol (PVA), amino acids, amines, surfactants, aryl pentafluoro compounds, poly(vinylpyrrolidinone) or polyethyleneimine etc.; polymers well known as liquid chromatographic (C_2, C_8, C_{18}) or gas chromatographic stationary phases (polyethylene glycol (PEG), phenylmethyl silicone) can be used for this purpose.

Due to the minimal total volume (μl) of the capillary the injection volume must not exceed few tens of nl (1–2% of the total volume), if a high separation efficiency is to be maintained in the system. The injection of these minute amounts of liquids poses problems of accuracy and precision, but modern instruments can assure reproducibility better than 2%. With the use of internal standards an even better precision can easily be achieved.

Modern instrumentation uses two injection principles: hydrodynamic or electrokinetic. In the first case, a pressure difference is generated between the two ends of the capillary, while the injection end is dipped in the specimen vial. Alternatively, a voltage difference is established between the sample's and the opposite vial, while samples are being injected. Although hydrodynamic injection (pressure driven) is nonselective (i.e. what is injected is representative of the sample composition), electrokinetic injection (potential driven) is selective. This is because the sample components enter the capillary according to their electrophoretic mobility, the mobility and concentration of total ions in the sample and the EOF (see in the next section). Electrokinetic injection allows field-amplified sample stacking, i.e. a highly efficient method for increasing analytical sensitivity (in terms of concentration), to be accomplished.

The high-voltage power supplies used for CE are generally able to give voltages up to 20–30 kV and currents up to 200–300 μA. Separations are generally carried out under constant potential, but voltage gradients or steps; and constant current separation is sometimes used. Because EOF in fused silica capillaries is usually directed towards the cathode, the common polarity in a capillary electropherograph is with the anode at the injection end of the capillary and the cathode at the opposite end, close to the detector; however, under specific analytical conditions, the polarity is reversed.

CE detection, most often with UV–visible radiation, is carried out in-capillary, in order to avoid any possible postseparation added volumes. This can

cause unacceptable band spreading (Fig. 1). Indeed, the wall of fused silica capillaries, after removing the polyimide coating, is highly transparent. Thus, picogram amounts of analytes can easily be detected even by UV absorption, but, due to the limitations in the sample volume (nl) which can be injected, the concentration sensitivity with UV detectors is limited to 10^{-6}M. However, this limit in sensitivity can be overcome by using other detection techniques, e.g. laser-induced fluorescence (sensitivity up to 10^{-12}M) or electrochemical detection (sensitivity up to 10^{-8}M) and/or by adopting high efficiency sample stacking methods, which can assure improvements in sensitivity of 2–3 orders of magnitude.

In current instrumentation, detection is most often based on UV(–visible) absorption; implementation of diode-array or fast-scanning UV detectors allows on-line recording of the UV spectra of the peaks, improving the information content of the CE analysis. Fluorescence (mostly, laser induced), electrochemical detection (conductimetric) and mass spectrometric (generally with electrospray ionization) detection modes have recently become commercially available. Other detection modes (including amperometric detection, thermooptical detection, chemiluminescence etc.) are still under development or need special instrumental arrangements.

A detection mode which has gained popularity in CE is the so-called 'indirect detection', which allows the determination of ionic molecules, such as small inorganic and organic ions, that are not directly detectable by the used CE detectors (e.g. do not absorb UV, are not fluorescent). Indirect detection is based on the addition to the running buffer of a ionic additive which is detectable at trace levels by the detector. This ionic compound, having the same charge and similar mobility to the analyte(s) of interest, is displaced from the zones where the analyte(s) migrate to preserve electroneutrality, and this displacement gives rise to 'reversed peaks', the area of which is proportional to the concentration of the given analyte(s). Drawbacks of this detection mode are sensitivity, which is usually lower than with the corresponding 'direct' mode, and the narrower range of linearity.

Separation Techniques

One of the most interesting characteristics of CE is that with the same instrumental hardware separations based on different physicochemical principles can be carried out. The most common CE separation techniques include: capillary zone electrophoresis (CZE), micellar electrokinetic capillary chromatography (MECC or MEKC), capillary electrochromatography (CEC), capillary isoelectric focusing (CIEF), capillary gel electrophoresis (CGE), capillary isotachophoresis (CITP) and capillary electrophoretic immunoassay (CEIA), which can be accomplished by simply changing the composition of the running buffer and the type of capillary.

Capillary zone electrophoresis (CZE)

CZE is a free-solution high-voltage electrophoresis in electrolyte-filled capillaries. Separation is based on the different electrophoretic mobility of analytes (μ_e), which under an electric field (E) migrate as sharp zones at different velocity ($v = \mu_e E$).

μ_e is described by the following equation:

$$\mu_e = \frac{q}{6\pi\eta r}$$

where q = ion charge, r = ion radius, η = solution viscosity.

For weak acids and bases, the charge and effective mobility (μ_{eff}) is dependent on the pK values of the different ionizable groups and on the pH of the running buffer.

In particular, for a weak acid

$$\mu_{eff} = \frac{K_a}{K_a + [H^+]}\mu_e^-$$

and for a weak base

$$\mu_{eff} = \frac{[H^+]}{K_a + [H^+]}\mu_e^+$$

Besides electrophoretic migration, a fundamental factor in CZE is electroendoosmosis. EOF produces an electrically driven flow of liquid, which in fused silica capillaries is directed from the anode to the cathode.

The linear velocity of the electroosmotic flow is described by the following equation:

$$v_{eo} = \frac{\varepsilon \ \zeta E}{4\pi\eta}$$

where ε is dielectric constant, ζ is zeta potential, η is solution viscosity and E is electric field.

The usual arrangement of a capillary electropherograph is such that injection is at the anodic end and detection is close to the cathodic end of the capillary. Considering also that EOF is generally oriented towards the cathode and that it is greater than the electrophoretic velocity of most analytes, it follows that cations, neutral species and anions will reach the detector in this sequence.

In fact, the migration velocity of cations and anions will result from the algebraic sum of their electrophoretic velocity and that of EOF. All neutral species

migrate at the same velocity as the EOF and consequently can not be resolved.

In the presence of the EOF, the migration velocity of the analytes follows the equation:

$$v = \frac{(\mu_{eo} + \mu_e)V}{L}$$

where μ_{eo} and μ_e are mobilities of EOF and of the analyte, respectively, V is potential and L is capillary length.

In CZE, the most important experimental factors controlling the separation are applied voltage, capillary length, buffer pH and composition, additives and capillary wall modifications.

The ideal CZE background buffer should have good buffering capacity at the chosen pH, low conductivity (to let high potentials act without unbearable heat production), buffer ion mobility matched to that of the analyte (to avoid peak distortion), and negligible interference with the used detection technique (e.g. low background UV absorbance at the adopted UV wavelength).

In practise, high buffer concentrations have been shown to reduce the wall-to-analyte interactions and, for a better buffering capacity, relatively high molarity background buffers are preferred in CZE, compatible with the resulting currents.

For most purposes, phosphate, borate, citrate and phosphate–borate buffers can be used. Zwitterionic buffers (Tris, 3-[(cholamidopropyl)dimethylammonio]-1-propane-sulfonate (CHAPS) etc.) are sometimes necessary in order to work at high concentrations without excessive currents (and consequently too high temperatures in the capillary generating loss of efficiency).

Buffer additives can greatly affect selectivity and resolution, introducing in the electrophoretic separation process new interactions and consequently additional selectivity. The most important additives in CE include: organic solvents (e.g. methanol, acetonitrile, isopropanol, tetrahydrofuran), anionic (e.g. sodium dodecyl sulfate (SDS)), cationic (e.g. cetyltrimethylammonium bromide (CTAB)) or neutral surfactants, organic amines, metal ions, urea, linear polymers (PEG, polyacrylamide, alkylated cellulose), complexing agents. Finally, if the nature of interactions between additives and analytes are stereoselective (with for example bile salts, cyclodextrins), a chiral selectivity can be achieved by the CE separation.

Micellar electrokinetic capillary chromatography (MECC)

As discussed above, in CZE, the charged molecules undergo electrophoretic separation, whereas all the neutral substances are pushed by the EOF to move at the same velocity towards the detector and, consequently, are not separated.

Around the middle of the 1980s, micellar electrokinetic capillary chromatography (MECC) (known also as MECK) was developed to resolve uncharged compounds by Prof. S. Terabe.

In this separation technique, an ionic surfactant forming micelles is added to the buffer and interacts with the solutes according to partitioning mechanisms. In this quasi-chromatographic system, EOF acts as the chromatographic 'mobile phase' pushing analytes (and micelles) towards the detector. The micelles represent the 'stationary phase' (better called 'pseudostationary'), and, because of their charge, tend to migrate backwards and selectively retard the analytes they interact with.

In a very common MECC system, the micellar phase is composed of the anionic surfactant SDS and the resulting anionic micelles are electrostatically driven towards the anode. However, because of the prevalence of EOF (running buffer is basic and generates a strong EOF directed towards the cathode), moving in the opposite direction, the resulting micelle migration is in the direction of the detector (close to the cathode). Consequently, the micelles selectively affect the 'mobility' of nonionic solutes (which otherwise equals that of EOF) they interact with. A selectivity based on partitioning of solutes between buffer and the lipophilic core of the micelles is thus introduced.

For ionizable/ionized solutes, separation mechanisms include electrostatic interactions between the ionic forms of the analytes and the charged surface of the micelles (ion-exchange like).

Selectivity changes depend on the nature of the micelles. Besides SDS and other anionic surfactants, other surfactants commonly used in MECC are hydrophobic-chain quaternary ammonium salts. The latter invert the EOF and consequently separations require reversed polarity.

Other factors crucial for the fine tuning of selectivity, such as organic solvents (methanol, isopropanol, acetonitrile, tetrahydrofuran), may be added to the separation buffer, in a similar way to reversed-phase chromatography. Also, urea (2–6 M) is known to increase the solubility in water of hydrophobic molecules and, in MECC, the separation of very lipophilic compounds is reported to be improved by highly concentrated urea.

Microemulsion capillary electrophoresis (MCE), using oil-in-buffer microemulsions instead of surfactant micelles, exerts a separation mechanism similar to MECC, but, so far, it has found limited application and is still mostly at an experimental level.

Capillary electrochromatography (CEC)

In CEC, a chromatographic stationary phase is contained in the capillary (packed or wall-immobilized) interacting with the solutes according to the usual chromatographic separation mechanisms. The mobile CEC phase is driven through the capillary by electroosmosis, not by pressure, as occurs in chromatography. This allows high resolving power and efficiency.

CEC is still at an early stage of development but looks particularly promising for separation of neutral hydrophobic molecules and in view of its coupling with MS.

Capillary isotachophoresis (CITP)

CITP is the capillary counterpart of classical isotachophoresis. In CITP, the ionic compounds migrate in discrete zones, at the same velocity, between a leading and a terminating electrolyte, have the highest and the lowest mobility of all analytes, respectively. CITP principles may be applied not only for separation, but also for sample pretreatment, before CZE, achieving concentration factors in excess of 100 times.

Capillary gel electrophoresis (CGE)

Gel electrophoresis is still the standard method for the separation and characterization of proteins (SDS–polyacrylamide gel electrophoresis (PAGE)), DNA fragment mapping and DNA sequencing. CGE is its instrumental evolution.

CGE capillaries may contain chemically crosslinked gels or noncrosslinked linear polymer matrices. Although they have higher efficiency and resolving power, capillaries filled with crosslinked gels are very delicate and prone to clogging. An alternative is represented by noncrosslinked sieving gels, generated by the physical entanglement of linear polymers (alkylated celluloses or linear polyacrylamide) dissolved at suitable concentrations in the running buffer. In effect, the noncrosslinked polymers exert a sieving mechanism similar to traditional crosslinked gels, but remain fluid and can be replaced by refilling the capillary by pressure application.

Capillary isoelectric focusing (CIEF)

CIEF is the replication in the capillary of slab gel isoelectric focusing, a popular separation mode in protein separation. In the classical isoelectric focusing, pH gradients are formed by letting a myriad of amphoteric compounds (ampholytes, the industrial product of the chemical breakdown of acrylic acid) arrange themselves side-by-side according to their isoelectric point (pI at which the net surface charge is zero), as long as a stable (constant voltage) electric field is established. At equilibrium, proteins are placed in the gel and are conveyed to the pH point corresponding to their pI (at isoelectric point, charge and mobility are null and there is no further migration). The same principles apply to CIEF, with the notable differences that separation is carried out in free solution, and that, after focusing, the various isoelectric bands have to be mobilized to the detector by hydrodynamic or electroosmotic methods.

Chiral separations

The separation of chiral compounds is gaining increasing attention in pharmacological/pharmaceutical science as well as in forensic toxicology. In reality, CE most often allows chiral separation to occur directly in the liquid phase, without the need of derivatization or of chiral stationary phases. The chiral selectors used in CE include cyclodextrins (CDs), Cu(II)-L-histidine, Cu(II)-aspartame, bile salts, crown ethers, proteins (bovine serum albumin, α_1-acid glycoprotein). Chiral resolution results from stereospecific interactions of the chiral selector with the two enantiomers giving rise to a difference in mobility between the two entities.

Capillary electrophoretic immunoassays (CEIA)

Coupling competitive immunoassays with CE-driven bound/free fraction separation has recently gained considerable attention, particularly whenever fluorescent tracers are used. Both CZE or MECC can be used to separate the bound from the free fraction of the tracer in a competitive immunoassay, allowing indirect but highly sensitive quantitation of several analytes in a variety of biological samples. Additionally, since different tracers can be easily spotted in one electropherogram, multicomponent immunoassays can be simultaneously carried out in the same experiment. Many commercial immunoassay reagents can be adapted to CEIA.

Applications

Illicit drug analysis and forensic toxicology

Since its start, CE has been applied to the analysis of drugs and pharmaceuticals and in pharmaceutical research and development. Although much later, CE has also been proposed as a tool for illicit drug analysis in both clandestine preparations and biological samples.

The first application of MECC to the analysis of illicit drug substances achieved the separation of a broad spectrum of compounds of forensic interest, including psilocybin, morphine, phenobarbital, psilocin, codeine, methaqualone, lysergic acid diethylamide (LSD), heroin, amphetamine,

chlordiazepoxide, cocaine, methamphetamine, lorazepam, diazepam, fentanyl, phencyclidine, cannabidiol and tetrahydrocannabinol. The protocol implied the use of 50 μm i.d. bare fused silica capillaries (length: 25–100 cm) using a buffer consisting of 8.5 mM phosphate, 8.5 mM borate, pH 8.5, 85 mM SDS and 15% acetonitrile; the applied voltage was 25–30 kV and detection was by UV absorption at 210 nm. Fluorescence detection (excitation wavelength 257 nm, emission 400 nm) was also successfully tested for fluorescent analytes (containing phenanthrene rings). Highly efficient separations of heroin acidic and neutral impurities, degradation products and adulterants have been obtained from seized heroin samples, and from street cocaine samples, the simultaneous resolution of benzoylecgonine cocaine, as well as *cis*- and *trans*-cinnamoylcocaine have been obtained.

In this first application, the analytical precision was characterized by a relative standard deviation (RSD) of about 0.5% for migration times and 4–8% for peak areas and peak heights (although peaks with migration times >40 min had worse precision). This phenomenon was ascribed to a rapid 'aging' of the buffer due to evaporation of organic solvent and water electrolysis. Frequent changes of the buffer and the use of electrolyte solutions with high buffering capacity was recommended to improve reproducibility. Additionally, since MECC and reversed-phase HPLC yielded uncorrelated elution patterns, they could be used as mutual countercheck procedures.

MECC, under similar conditions, has been used in combination with a fast-scanning UV detector for the analysis of illicit preparations of heroin and cocaine resulting in the identification of the individual peaks on the basis of both migration times and UV spectra.

Several other authors have confirmed the suitability of MECC with either anionic or cationic micelles to analyze clandestine preparations of illicit drugs.

Comparison of MECC with HPLC and GC showed good quantitative correlation between the different techniques. Precision was slightly worse than with HPLC, but the resolution of complex samples was better with CE. Also, CE proved highly reliable in interlaboratory proficiency tests for illicit heroin and cocaine analysis.

In drug screening, CZE proved to be a realistic alternative to MECC when using plain 50 mM phosphate buffer, pH 2.35, in a 75 μm i.d. (60 cm long) bare fused silica capillary. Detection was by UV absorption at 214 nm. In only 11 min, the separation of 17 basic drugs of forensic interest including methapyrilene, brompheniramine, amphetamine, methamphetamine, procaine, tetrahydrozoline, phenmetrazine, butacaine, medazepam, lidocaine, codeine, acepromazine, meclizine, diazepam, doxapram, benzocaine and methaqualone was achieved. Drugs with lower pK_a values and consequently less positive charge, showed higher migration times, but other factors (such as molecular size, tendency to interact with the column and ability to form doubly charged species) affected electrophoretic mobility, allowing the resolution of molecules with the same charge-to-mass ratio. RSDs of migration times were <1%, whereas peak-area RSDs ranged from 1.5 to 4.3% (worse reproducibility was found for analytes with pK_a values close to the pH of the background buffer and, consequently, with very slow migration). CZE may have advantages over MECC for drug screening and, particularly, simple background electrolyte preparation and shorter analysis times. The main limitation is the inability to analyze acidic, neutral and basic drugs simultaneously, as is possible with MECC.

The CZE mobility of as many as 550 basic (analyzed in 100 mM phosphate, pH 2.38) and 100 acidic drugs (analyzed at pH 8.50) has been determined for the rapid screening of blood samples.

The complementary nature of CZE and MECC for the identification of 17 illicit drugs and related compounds has been clearly demonstrated in a study in which MECC with SDS at pH 9.2 gave highly noncorrelated separation with CZE at pH 2.35.

Chiral discrimination is of primary importance in the analysis of amphetamine substances for investigations on the synthetic methodologies and CE is the most effective analytical tool for this purpose. In effect, it offers high efficiency chiral resolution, high speed and flexibility, low costs, and no need of sample derivatization. After preliminary attempts by using derivatization with a chiral reagent (2,3,4,6-tetra-*O*-acetyl-beta-D-glucopyranosyl isothiocyanate) to form diastereo-isomers, followed by MECC, direct chiral analysis has been introduced by using native or derivatized β-CDs and this is now a standard methodology for the separation of enantiomers of amphetamine/methamphetamine, including ring substituted analogues.

CE was introduced in the field of forensic and clinical toxicology with a MECC method based on a phosphate–borate buffer, pH 9.1, with 75 mM SDS for the qualitative determination in urine of drugs of abuse (and their metabolites) including benzoylecgonine, morphine, heroin, 6-monoacetylmorphine (MAM), methamphetamine, codeine, amphetamine, cocaine, methadone, methaqualone and some major benzodiazepines.

Sample extraction was carried out by using commercially available 'double mechanism' (cation exchange and reversed-phase) solid-phase extraction

cartridges widely adopted in GC and HPLC. The extract from 5 ml of urine was dried and the residue redissolved with 100 μl of running buffer and injected. Under these conditions, a detection limit of $100\,\mathrm{ng\,ml^{-1}}$ was easily achieved, meeting the required sensitivity for confirmation of current immunoassays. Peak identification was also achieved by comparison of the on-line recorded UV spectra of the peaks with computer-stored models of known compounds.

Similar MECC approaches, based on separations in phosphate–borate/SDS buffers and multiwavelength UV detection, have been proposed for the determination of many drugs of abuse in biological fluids, in some instances without any sample pretreatment. However, CZE also proved suitable for this purpose.

Table 1 summarizes some paradigmatic methods for the determination of a wide range of abused substances.

Although drugs administed at high doses, such as

Table 1 Selected CE methods for analysis of drugs of forensic interest in biological fluids

Analyte	*Sample/sample preparation*	*Method*
Benzoylecgonine, morphine, heroin, 6-MAM, codeine, methamphetamine, amphetamine, cocaine, methadone, methaqualone, benzodiazepines	Urine/SPE	MECC: uncoated fused silica capillary, phosphate–borate buffer pH 9.1 with 75 mM SDS; detection: UV
Barbital, allobarbital, phenobarbital, butalbital, thiopental, amobarbital, pentobarbital	Urine/SPE Serum/direct	MECC: uncoated fused silica capillary, phosphate–borate buffer pH 7.8 with 50 mM SDS; detection: UV
11-nor-delta-9-tetrahydrocannabinol-9-carboxylic acid (THC metabolite)	Urine/SPE, hydrolysis	MECC: uncoated fused silica capillary, phosphate–borate buffer pH 9.1 with 75 mM SDS; detection: UV
Morphine-3-glucuronide	Urine/SPE	MECC: uncoated fused silica capillary, phosphate–borate buffer pH 9.8 with 75 mM SDS; detection: UV CZE: uncoated fused silica capillary, phosphate–borate buffer pH 9.8:
Ephedrine, norephedrine	Urine/direct	CZE: uncoated fused silica capillary, 50 mM phosphate buffer pH 9.5, 1% acetonitrile; detection: UV
Acebutolol, alprenolol, atenolol, labetalol, metoprolol, nadolol, timolol oxprenolol, pindolol, propranolol,	Urine/direct or SPE	MECC: uncoated fused silica capillary, 80 mM phosphate buffer pH 7 with 10 mM CTAB; detection: UV
Methapyrilene, brompheniramine, codeine, lidocaine, amphetamine, methamphetamine, meclizine, procaine, tetrahydrozoline, phenmetrazine, butacaine, medazepam, acepromazine, diazepam, doxapram, benzocaine, methaqualone	Urine/LLE Serum/LLE	CZE: uncoated fused silica capillary, 50 mM phosphate buffer pH 2.35; detection: UV
Flunitrazepam, diazepam, midazolam, clonazepam, bromazepam, temazepam, oxazepam, lorazepam	Urine/SPE	MECC: uncoated fused silica capillary, phosphate–borate buffer pH 9.3 with 75 mM SDS; detection: UV
Methadone, EDDP (urinary metabolite)	Urine/SPE	CZE: uncoated fused silica capillary, 50 mM borate buffer pH 8.9; detection: UV
Clenbuterol	Urine/LLE	CZE: uncoated fused silica capillary, 25 mM citrate buffer pH 5; detection: UV
26 tricyclic drugs, including phenotiazine, imipramine, promazine, amitriptyline, clozapine, amiodarone, prochlorperazine, chlorpromazine, etc.	Urine/LLE	MECC: uncoated fused silica capillary, 40 mM borate buffer pH 9.5 with 10 mM Na taurodeoxycholate; detection UV
Pholcodine, 6-MAM, morphine, heroin, codeine, dihydrocodeine	Urine/SPE	CZE: uncoated fused silica capillary, 100 mM phosphate buffer pH 6.0; detection: UV
LSD, nor-LSD, iso-LSD, iso-nor-LSD	Blood/LLE	CZE: uncoated fused silica capillary, 250 mM citrate/acetate buffer pH 4.0-methanol, 30:70; detection: LIF
Amphetamine, methamphetamine and related compounds	Urine/LLE-SPE Derivatization	CZE: uncoated fused silica capillary, 50 mM acetate buffer pH 3.0; MECC: 50 mM borate buffer pH 9.3 with 15 mM SDS; detection UV-LIF

Abbreviations: CTAB: *N*-cetyl-*N*,*N*,*N*-trimethylammonium bromide; CZE: capillary zone electrophoresis; LIF: laser-induced fluorescence detection; LLE: liquid/liquid extraction; MECC: micellar electrokinetic capillary chromatography; SPE: solid phase extraction; SDS: sodium dodecylsulfate.

barbiturates, may be determined directly by CE, the relatively poor sensitivity in terms of concentration of this technique most often necessitates that the sample is extracted and concentrated before injection. A stepwise solid-phase extraction of the most common drugs of abuse from human urine, preliminary to MECC, used commercial cartridges exhibiting hydrophobic and ion-exchange interactions and was reported to give 'clean' electropherograms even with 50-fold concentrated urine. It has also been observed that MECC with SDS in plain aqueous buffers failed to resolve some analytes (e.g. amphetamine, methamphetamine, methadone and benzodiazepines) and that it could be obtained by adding low percentages of acetonitrile (5–10%).

Hair analysis is an innovative field of forensic toxicology, shedding light on former, chronic exposure to illicit drugs. The usual analytical strategy is based on hair extraction, extract purification, immunological drug screening and chromatographic (GC, GC–MS, HPLC) confirmation. To this aim, CE offers advantages in terms of peculiar separation mechanism and minimal need of hair sample.

By using a CZE method based on an uncoated fused silica capillary, a background buffer composed of 50 mM borate, pH 9.2 and UV detection, the simultaneous detection of cocaine (238 nm) and morphine (214 nm) has been accomplished. The use of sample stacking techniques allowed a >100-fold increase in sensitivity (**Fig. 2**), with the possibility of recording UV spectra of the drugs in hair.

Due to the fundamental role of mass-spectrometry (MS) in modern analytical toxicology, great attention is at present being paid to the CE–MS coupling. The development of on-line CE–MS started in the late 1980s and more recently commercial instruments with electrospray interfaces have appeared (CE-ESI-MS). CE-MS has been introduced in the pharmaceutical field and to some extent in forensic toxicology.

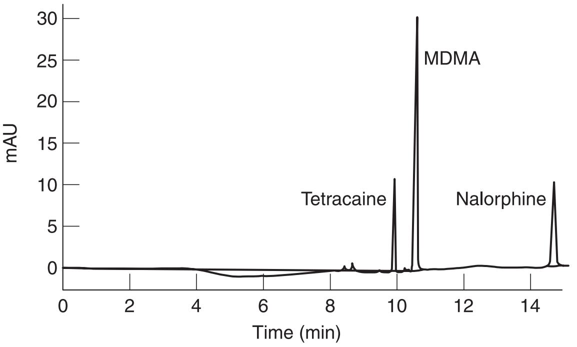


Figure 2 Electropherogram of an extract from a sample of hair from a user of 'ecstasy' containing 3,4-methylenedioxymethamphetamine (MDMA) at a concentration of 4.0 ng mg^{-1}. Tetracaine and nalorphine were added as internal standards. Injection was by electromigration under field amplified sample stacking conditions. Separation was by CZE using 100 mM phosphate pH 2.5 as the running buffer. Detection was by UV absorbance at 200 nm. (Reprinted with permission from Tagliaro F, Manetto G, Crivellente F, Scarcerlla D and Marigo M (1998) Hair analysis for abused drugs by capillary zone electrophoresis with field-amplified sample stacking. *Forensic Science International* **92**: 201–211.

Analysis of gunshot residues and explosive constituents

The analysis of gunshot and explosive residues, for intelligence or courtroom purposes, is a fundamental area of forensic science.

Both inorganic and organic constituents are of crucial interest and are currently analyzed by the most sophisticated techniques, including scanning electron microscopy neutron activation analysis, mass spectrometry, X-ray and infrared techniques, and all types of chromatography. CE can complement more established methods, especially in the analysis of organic and inorganic compounds.

In this field, the first contributions date back to the early 1990s, when MECC was used for the separation and determination of the major organic constituents in reloading powders, explosive materials and in gunshot residues. The separation method used uncoated fused silica capillaries and a buffer composed of 2.5 mM borate, 25 mM SDS; detection wavelength was 250 nm (200 nm for nitroglycerin). As many as 11 components of gunshot residues (including nitroguanidine, nitroglycerin, 2,4-dinitrotoluene (DNT), 2,6-DNT, 3,4-DNT, 2,3-DNT, diphenylamine (DPA). *N*-nitrosoDPA, 2-nitroDPA, ethylcentralite and dibutyl phtalate) and 15 high-explosive constituents (including nitroguanidine, ethylene glycol dinitrate, diethylene glycol dinitrate, 1,3,5-trinitro-1,3,5-triazacyclohexane (RDX), nitroglycerin, 2,4,6-trinitrotoluene (TNT), pentaerythritol tetranitrate, picric acid) were successfully resolved in 10 min. In addition, since several analytes have characteristic UV spectra, multiwavelength absorbance analysis helped their identification. Application to forensic cases included the investigation of spent ammunition casings, reloading powders and plastic explosives (Semtex, C4, Detasheet, Tovex).

Specific sample collection procedures for the CE analysis of gunshot residues have been developed, based on the use of masking adhesive tape (1 inch square sections). The film lifts were examined with a binocular stereoscope and the gunshot residue particles collected with tweezers and analyzed by MECC. This adhesive film collection was found to be clearly superior to the traditional swabbing. By the combined use of improved collection and microsample analysis, determination was possible of individual

particles collected at the skin surface. Firing range experiments with subjects trying different weapons demonstrated that characteristic gunshot residue constituents could be recovered by adhesive film lifts from both the weapons and the hands (**Fig. 3**).

In the case of home-made explosives (the so called low-explosives), anions and cations left behind from the blast represent useful pieces of evidence to determine the type and source of the explosive mixture used. Moreover, traces of the explosive mixture can be looked for in the environment where the device was assembled and/or on the perpetrator.

Although ion chromatography is the traditional tool for this purpose, CE can be considered a useful complementary method. To analyze postblast anions, CZE was used in borate buffer (2 mM borate, 40 mM boric acid) with 1 mM diethylenetriamine as EOF modifier at a final pH of 7.8 and a dichromate chromophore (1.8 mM); detection was indirect by UV absorbance at 280 nm. By using this method, chloride, nitrite, nitrate, sulfate, sulfide, chlorate, carbonate, hydrogencarbonate, cyanate, thiocyanate and perchlorate could be determined in a single run. A comparison of CZE with ion chromatography in real cases showed major differences between the two separation patterns, allowing mutual confirmation of the results. The reproducibility of the migration times in CZE using the same capillary was characterized by day-to-day RSD better than 1%, and could be further improved by adding bromide as a marker. Practical tests were carried out on pipe-bomb fragments experimentally prepared with different explosive mixtures (potassium chlorate–vaseline; black powder; smokeless powder; a mixture of black and smokeless powder) and detonated. Fragments from bombs were simply extracted with water, filtered and analyzed.

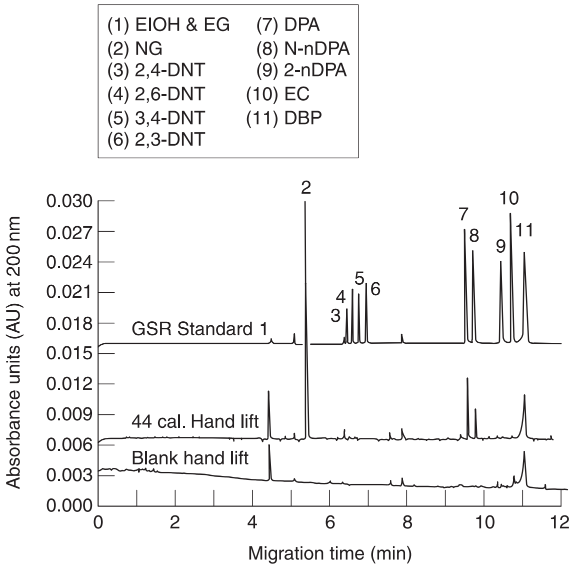


Figure 3 Electropherograms from MECC analyses of (top) a solution of gunshot residue standards, (middle) an extract of a single gunshot residue particle on a film lift from a hand which fired a 44 caliber revolver and (bottom) an extract from a film lift blank. Buffer: 25 mM sodium dodecyl sulfate, 2.5 mM borate, pH 8.5; voltage: 25 kV. (Reprinted with permission from Northrop and MacCrehan (1992) *J. Liquid Chromatography* 15(6):1061–1063.)

A similar CZE approach has also been used to determine inorganic anions and cations, such as nitrate, chlorate and ammonium as postblast residues of chlorate or emulsion explosives.

Analysis of pen inks

In forensic laboratories, TLC, column chromatography and slab gel electrophoresis have been used to investigate ink composition, but, again, CE has been applied with very encouraging results.

The water-soluble components of black inks from 26 marking pen specimens were analyzed by CZE using 15 mM borate buffer, pH 8.8, uncoated capillary and UV detection at 214 nm. CZE at basic pH (100 mM borate, pH 8.0, with 20% methanol) has also been used with diode array UV and LIF detection, to reliably classify almost all the components of 17 blue and black ink samples from different manufacturers. However, only initial results have been reported from dried paper-extracted inks.

Separation of biopolymers

Proteins CE has been widely used for peptide and protein analysis, but so far very little in the forensic field.

A subject of crucial interest in forensic (and clinical) medicine is the objective diagnosis of chronic alcohol abuse.

Although acute ethanol intoxication can be diagnosed by measuring blood concentrations, objective diagnosis of chronic/subchronic excessive intake is still an open problem. To this aim, several methods have been proposed including ethanol metabolites and congeners, enzymes, acetaldehyde adducts, high density lipoprotein cholesterol, 5-hydroxytryptophol and 5-hydroxytryptophol-3-acetic acid, dolichols etc. Among these, carbohydrate-deficient transferrin (CDT) is considered the most reliable objective piece of evidence of continuous excessive alcohol consumption.

The basis of CDT is a microheterogeneity of serum transferrin, the major iron transporting glycoprotein, which may contain from 0–8 sialic acid residues. The major human transferrin isoform (>90%) contains four sialic acid residues, but, as a result of alcohol abuse, less glycosylated isoforms (disialo- and asialo-

transferrin) are reported to increase and are collectively named CDT. The exact mechanism of this phenomenon has not yet been clarified, but, on the basis of a large body of literature, it is believed that an ethanol intake >50–$80\,\mathrm{g\,day^{-1}}$ for one to two weeks leads to an abnormal increase of CDT, which has a half-life of about 15 days. Diagnostic specificity, a crucial parameter for CDT application in the forensic environment, is reported to be good-to-excellent (90–100%) with a sensitivity of 50–90%.

The current determination of CDT is based on immunoassays coupled to sample pretreatment using ion-exchange cartridges. More sophisticated methods of analysis include isoelectric focusing combined with immunofixation, zone immunoelectrophoresis or Western blotting, anion-exchange chromatography or chromatofocusing followed by immunoassays and HPLC with UV detection.

CE analysis of CDT has been optimized and validated in clinical cases and its application to a forensic context is today a realistic option.

A simple CZE method using 100 mM borate buffer pH 8.3 and uncoated capillaries (20 μm i.d.) with UV detection at 200 nm wavelength proved able to separate and determine quantitatively disialo- and trisialotransferrin isoforms (i.e. the glycoforms of major clinical relevance) in human serum within a reasonable time (20 min) and without the need for complex sample pretreatment (limited just to iron saturation and 1/10 dilution in water).

Other proteins of forensic interest which are susceptible of CE analysis include human protein variants used as phenotypic markers for identification, hemoglobins and saliva and semen proteins.

DNA fingerprinting The crucial importance of DNA analysis in modern forensic science (personal identification, paternity testing) need not be further emphasized here. However, CE is increasingly being applied in the field, particularly since the introduction of fully automated multicapillary instrumentation dedicated to nucleic acids analysis (polymerase chain reaction (PCR) amplicons and DNA sequencing).

The analysis of DNA polymorphisms is today largely focused on length polymorphisms and sequence variation. Length polymorphisms are by far the most important class of genetic variability used to identify humans. Electrophoresis on agarose or polyacrylamide slab gels is the traditional procedure for determining individual genotypes. Generally, short (<1000 base pairs long: amplified fragment length polymorphisms, AmpFLPs) or very short (down to 200 base pairs; short tandem repeats, STRs) products of PCR are targeted, differing by as little as 1% of the overall length.

In a very competitive field, CE has been shown to give as fast, reproducible and reliable DNA separation as polyacrylamide and agarose slabs. As traditional gel electrophoresis, the CE separation mechanism is that of molecular sieving (CGE).

To achieve length polymorphism separation, crosslinked gel-filled columns have been used (typically 2–6% T and 3–6% C polyacrylamide). However, gel-filled columns generally have a short lifetime because of localized overheating, irreversible binding of DNA strands to gel and capillary clogging. As an alternative, noncrosslinked gel sieving media have been introduced, based on water solutions of linear polymers, such as methylcellulose or linear polyacrylamide, at a suitable concentration. In these media, the mutual entanglement of target and matrix molecules has a sieving effect and leads to separation by fragment length.

Fresh noncrosslinked refills are available by simply flushing new buffer into the capillary. On the other hand, noncrosslinked gel sieving systems usually show a weaker resolving power and lower efficiency, compared to crosslinked gel capillaries.

Noncrosslinked gels in CGE systems have been applied to problems of forensic genetics. The sieving buffer system was based on 0.5% (w/v) hydroxyethylcellulose in 100 mM Trisma base and 100 mM boric acid, adjusted to pH 8.7 with cesium hydroxide; the intercalating agent ethidium bromide and EDTA were added at suitable concentrations. Phenylmethyl-coated silica capillaries (70 cm × 100 μm i.d.) were used.

Few but important specific problems arise from the coupling of CE and PCR. In short, the presence of inorganic ions at high concentrations and excess primers in the PCR mixture has a detrimental effect on the electrokinetic injection of DNA molecules, as the small inorganic anions are injected preferentially, owing to their more favorable mass/charge ratio. Therefore, the PCR samples have to be ultrafiltered and/or dialyzed. Furthermore, optimum polymer concentration is a crucial point. A concentration 0.5–1.75% has a positive effect on resolution of small fragments, likely reflecting a reduction in effective pore size, although other interactions cannot be excluded. DNA fragments in the range 150–400 base pairs may require even higher concentrations of the polymer, which, on the other hand, has a negative effect on the resolution of higher molecular mass fragments. Besides, the upper limit of polymer concentration is determined by the viscosity of the arising solution, as highly concentrated polymer solutions may not be pumped into the tiny separation capillary. Increasing ethidium bromide concentration from 0.635 to 6.35 mM, reportedly, improves resolu-

tion, but higher concentrations have a deleterious effect on the column performance, particularly if left for longer periods in the column.

Noncrosslinked gels have been used for sieving CE with UV detection for the analysis of two genetic markers, namely D1S80 (repeat unit of 16 base pairs) and SE 33 (repeat unit of 4 base pairs) by using 0.635–1.27 μM ethidium bromide in the background buffer containing 0.5% hydroxyethylcellulose.

By using a similar noncrosslinked gel CGE system, LIF detection to PCR amplified DNA has been applied. Since DNA has only a weak native fluorescence, PCR products must either be tagged using fluorescent primers, or stained with fluorescent dyes. The use of fluorescent dyes intercalating DNA is reportedly more sensitive than PCR tagging because a greater number of fluorescent molecules can be introduced. Also, DNA intercalation is known to improve the separation, by increasing the rigidity of the fragments. YO-PRO-1 was adopted for DNA intercalation because of its strong binding constant to DNA and little intrinsic fluorescence, when not bound to DNA. The application of noncrosslinked gel CGE with LIF detection to the analysis of polymerase chain reaction-amplified DNA fragments showed an excellent sensitivity, about 500-fold better than with UV detection. With this procedure, the alleles from a number of loci of interest in forensic genetic typing were investigated with success.

After these pioneering articles, numerous papers have been published on noncrosslinked CGE–LIF analysis of forensic DNA, particularly after the recent introduction of fully automated instrumentation for DNA fragment analysis and DNA sequencing. The most advanced approaches make use of multiwavelength fluorescence detection and/or multicapillary systems to increase reliability and productivity, respectively.

Conclusion

Discussing the admissibility of new science or science tools in the USA courts it has been stated:

> Given the economical and efficiency advantages of CE for forensic science and the judicial process, expert testimony based on CE analysis can be anticipated. The legal criteria of *Daubert*, as long as they are met by the scientific community, will allow CE and other legitimate scientific processes into evidence as acceptable expert testimony. (Kuffer *et al.* 1996)

In fact, the reputation of CE as a sound technology in several applied analytical sciences, including forensic science, is now well established.

Versatility (reflected in a variety of application fields) is probably the quintessential feature of CE, which is indeed day after day endorsing its image as not merely 'a new analytical tool', but a new 'dimension' in separation science. Of course, CE requires specific experience and new skills, which are still scarce in the forensic science environment, to the increasing cohort of new adepts.

Complementarity, the possibility of interfacing with other established techniques (spectroscopic, chromatographic, mass-spectrometric, electrophoretic) and the ability to analyze tiny amounts of samples are other outstanding features of CE of particular relevance in the forensic field.

On the other hand, CE has a tremendous innovation potential (e.g. immunochemical multianalyte drug assays, CE on a chip) that requires a really openminded, interdisciplinary approach, which is still rare in the quite conservative field of forensic science. Undoubtedly, in many areas of potential application of forensic CE, e.g. in forensic toxicology, the electrophoretic background of scientists is, generally speaking, poor, whereas other areas, such as forensic biology, have little familiarity with instrumental analysis.

For this reason, seminal activity on CE and personnel training are priorities for the near future.

In conclusion, we believe that, when forensic scientists will be more familiar with the new concepts of CE, this fascinating technique will become a 'normal' tool in the hands of forensic scientists for everyday work and will no longer be regarded as an exotic technique closer to academia than to the 'real world'.

See also: **Analytical Techniques**: Separation Techniques. **Deoxyribonucleic Acid**: Parentage Testing; Significance. **Clinical Forensic Medicine**: Evaluation of Gunshot Wounds. **Document Analysis**: Ink Analysis. **Drugs of Abuse**: Analysis. **Explosives**: Analysis. **Hair**: Comparison: Microscopic; Comparison: Other. **Analytical Techniques**: Capillary Electrophoresis in Forensic Biology.

Further Reading

Altria KD (1993) Capillary electrophoresis for pharmaceutical research and development. *LC-GC International* **6**:616–619.

Everaerts FM, Beckers JL and Verheggen TP (1976) *Isotachophoresis: Theory, Instrumentation and Applications*. Journal of Chromatography Library vol. **6**: Amsterdam: Elsevier.

Kuffner Jr CA, Marchi E, Morgado JM and Rubio CR (1996) Capillary electrophoresis and Daubert: time for admission. *Analytical Chemistry* **68**:241A–246A.

Li SFY (1994) *Capillary Electrophoresis, Principles, Practice and Applications*. Amsterdam: Elsevier.

McCord BR (1998) Capillary electrophoresis in forensic science. *Electrophoresis* **19**:1–124.
Northrop DM, McCord BR and Butler JM (1994) Forensic applications of capillary electrophoresis. *Journal of Capillary Electrophoresis* **1**:158–168.
Tagliaro F and Smith FP (1996) Forensic capillary electrophoresis. *Trends in Analytical Chemistry* **10**:513–525.
Tagliaro F and Pascali VL (1998) Capillary electrophoresis in the forensic sciences. *Forensic Science International* **92**:73–280.
Tagliaro F, Turrina S, Pisi P, Smith FP and Marigo M (1998) Determination of illicit and/or abused drugs and compounds of forensic interest in biosamples by capillary electrophoretic/electrokinetic methods. *Journal of Chromatography B* **713**:27–49.
Tagliaro F, Crivellente F, Manetto G, Puppi I, Deyl Z and Marigo M (1998) Optimized determination of carbohydrate-deficient transferrin isoforms in serum by capillary zone electrophoresis. *Electrophoresis* **19**:3033–3039.
Terabe S (1989) Electrokinetic chromatography: an interface between electrophoresis and chromatography. *Trends in Analytical Chemistry* **8**:129–134.
von Heeren F and Thormann W (1997) Capillary electrophoresis in clinical and forensic analysis. *Electrophoresis* **18**:2415–2426.

Gas Chromatography

V Cirimele, Institut de Médecine Légale, Strasbourg, France

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0784

Introduction

The term 'Forensic toxicology' covers any application of the science and study of poisons to the elucidation of questions that occur in judicial proceedings. All the techniques of analytical chemistry used by the forensic toxicologist, from color tests to mass spectrometry, are based on the molecular structure of the compounds involved, whereas toxicity is related to the dose. Toxicity of a poison is a biological concept, as anything can become a poison if it exceeds the threshold limit of the organism's ability to deal with it. If the poison is not specified by name, the request to 'test for poisons' is a major problem for the forensic toxicologist because there is no single chemical method of analysis capable of detecting all the various poisons. The forensic toxicologist needs a repertoire of standard methods that can be modified according to the nature of the investigation, the type and amount of material for analysis, and the time and resources available. Gas chromatography is one specific form of the more general separation process of chromatography.

Fundamentals

Gas chromatography (GC), like other forms of chromatography, is a method of separating mixtures of substances of analytical interest, either from each other or from an extraction residue. The separation is performed on a column containing the stationary phase, either solid or liquid, which is maintained at a defined temperature in an oven and at a defined flow of carrier gas (mobile phase). When a mixture of substances is injected at the inlet, each component is swept towards the detector and is partitioned between the stationary phase and the gas phase. Molecules with greatest affinity for the stationary phase spend more time in that phase and consequently take longer to reach the detector. The detector produces a signal dependent on the structure of substance passing through it. Each substance passing through the column will have a characteristic retention time, which is defined as the time from the injection to peak maximum.

Columns

The GC columns most widely used fall into two distinct categories: packed and wall-coated open tubular (WCOT) columns, commonly known as capillary columns because their inner diameters are small. Packed columns were developed first, but since about 1980 the commercial availability of highly efficient and rugged fused silica WCOT columns has resulted in their use dominating the gas chromatography coupled to mass spectrometry (GC–MS) field, especially for the analysis of trace amounts of specific organic compounds in complex mixtures.

Packed columns

Packed columns do not possess the high efficiencies and separating capabilities of WCOT columns, but they do have capacities which simplify sample introduction techniques and provide a larger quantity of sample component for introduction to the detector. Typical packed columns for analytical work are 2–5 m in length and have an internal diameter of 2 mm. They use carrier gas flows of 20–50 ml min^{-1}. The most common support particles are formed from diatomites, which are skeletons of single-celled algae.

These are prepared by molding diatomaceous earth into bricks and drying them in an oven. The bricks are then crushed and screened to particles in the 80–120 mesh size (80–120 openings per inch (2.54 cm) in a screen). These particles are very strong, with a high specific surface area (1–20 $m^2 g^{-1}$) and good pore structure (1–2 μm).

WCOT columns

WCOT columns were originated in 1956 when Golay first showed them to be theoretically ideal. They underwent major advances when pioneering work on high performance glass WCOT columns occurred in Europe between 1975 and 1980. These columns are now made of pure silica and are extremely rugged when an external coating of polyimide polymer is applied. The technology of producing high quality columns of controlled internal diameter and stationary film thickness has advanced considerably. Commercially available fused silica WCOT columns with polar and nonpolar stationary phases have a consistently high efficiency and give excellent analytical results. They are produced in lengths of 10–100 m with internal diameter of 0.20–0.35 mm and use carrier gas flow rates of 2–5 ml min^{-1}. Fused silica WCOT columns have the additional advantage over glass or metal columns of being physically flexible, so much so that they may actually be tied into a knot without breaking. To separate volatile compounds correctly, it is better to use a large stationary film thickness (alkanes C1 to C15 with 5 μm). In contrast, a lower stationary film thickness (0.12 μm) will optimize separation of C10 to C40.

Capillary columns offer other advantages over packed columns. Their superior resolution can be used to separate complex mixtures or to increase the certainty that a single compound is correctly identified. The high efficiency results in tall narrow peaks, which considerably enhance the signal-to-noise ratio and consequently the detection limits. Short columns (2–10 m) can be used to give similar resolution to packed columns, but in a shorter time.

Choice of the Stationary Phases

Over 700 substances have been used as stationary phases. GC may be divided into gas–solid chromatography (mainly adsorptive processes) and gas–liquid chromatography (mainly partition), depending on whether the stationary phase is a solid or a liquid at its operating temperature. If the stationary phase is a liquid, it must be coated on a support for packed column chromatography. For capillary column chromatography, the stationary phase may be coated directly on to the walls of the column, or on to a support which is bonded to the glass walls. The stationary liquid phase may be chosen from over 100 phases available, but in practice fewer than 10 are in common use (**Table 1**). Some examples of phases in gas–solid chromatography are given in **Table 2**.

In general, nonpolar compounds chromatograph best on nonpolar phases, and polar compounds on polar phases, but this is not necessarily the decisive factor. For example, alcohols, and particularly ethanol, are well separated on polar columns, such as polyethylene glycol. Nonpolar phases, like polydiphenyldimethyl siloxane, also give good results: separation is obtained using the volatilization temperature of the different constituents of the mixture after concentration of the analytes in the head column. In this case, a partition process is followed by an adsorptive one. Ethanol and 1-propanol are well separated (due to their different boiling points), in contrast with results using a polar stationary phase.

Optimization of Oven Temperature

For a particular separation, the lowest temperature compatible with a reasonable analysis time should be used. If the time is excessive, it is generally better to

Table 1 Some examples of phases in gas–liquid chromatography

Support materials	*Separated compounds*
Apiezon L (hydrocarbon grease)	Barbiturates, amphetamines
SE-30, OV-01, OV-101 (dimethyl silicone polymers)	Separation on the basis of molecular weight
Apolane-87 (high temperatue non-chiral hydrocarbon phase)	Many drugs
Carbowax 20 M (polyethylene glycol)	Alkaloids, basic drugs, amphetamines
OV-17 (phenyl methyl silicone, moderately polar silicone phase)	Many drugs
XE-60 (cyanoethyl silicone); OV-225 (cyanopropyl phenylmethyl silicone)	Steroids
Polyester	Fatty acid esters, barbiturates
Polyamides (Poly A 103)	Barbiturates, ternary amine tricyclic antidepressants
Chirasil-Val	Optical enantiomers (amino acids, polar drugs)
Mixed phases	Anticonvulsant drugs

Table 2 Some examples of phases in gas–solid chromatography

Stationary phase	*Separated compounds*
Molecular sieve (4A, 5A, 13X)	Inorganic gases, O_2, N_2, CO, in blood
Silica gel	Inorganic gases, CO_2, CO, H_2, N_2
Chromosorb/Porapak	Fatty acids, amines, alcohols
Tenax	Trapping of volatile substances
Carbopak B and C	Hydrocarbons C1 to C10, ethanol in blood, substances abused by 'glue sniffers', ethylene glycol in blood

reduce the stationary phase loading than to increase the column temperature.

In a screening procedure for complex mixtures with components of widely varying retention characteristics, it may be very difficult or impractical to select a column temperature that will allow all the components to be resolved. It is therefore necessary to vary the column temperature throughout the analysis, starting with a low temperature and finishing with a higher value. For mixtures with components of the same family, the oven temperature can start at a higher level and separation can also be obtained under isothermal conditions.

There is a maximum temperature at which a column can be operated, and there is also a minimum temperature below which the efficiency will drop sharply. The stationary phase must be a liquid at the temperature of operation, and if a column is run at too low a temperature to obtain longer retention times, the stationary phase may still be in the solid or semisolid form.

Gas Pressure and Flow Control

In order to perform accurate and reproducible GC, it is necessary to maintain a constant carrier gas flow. Under isothermal conditions, simple pressure control is adequate for packed or capillary columns. Flow control is highly desirable, if not essential, during temperature programming with packed columns and can be used to advantage with on-column injectors on capillary columns. Carrier gas flow should be optimized for a particular column and a particular carrier gas.

Introduction of Samples

Sample introduction in a GC analysis is critical. Poor introduction technique can reduce column resolution and the quality of quantitative results. The sample must be injected as a narrow band on to the head of the column and contain a composition truly representative of the original mixture.

Introduction of the sample to the column is the injector's function. Gases can be injected using a rotary valve containing a sample loop of known volume. Solid samples can be dissolved in a suitable solvent and converted to a vapor by the temperature at the inlet upon injection. The most common technique is liquid injection through a self-sealing septum into a heated injection port.

Alcohol can be introduced to packed columns by injecting a sample of blood directly diluted with water. The glass fiber insert will trap all nonvolatile compounds and the total flow of carrier gas will pass through the column. The injector temperature selected must be as low as possible to avoid column pollution or contamination. One limitation is the frequent change of insert necessary.

Split–splitless injection

Split and splitless injectors are more conventional today. For split injection, a flow of carrier gas will purge the septum and another will pass through the vaporization chamber. This later will be separated between column flow and purge. The ratio of these two flows (the split ratio) is the proportion of injected sample that reaches the column. The function of the splitter is not to reduce sample volume but to ensure that the sample enters the column as a plug and is not exponentially diluted, and to prevent overloading of the column with sample.

Although the split method of injection does prevent column overloading, the fraction which reaches the column may not be representative of the original sample because it is a flash vaporization technique: higher molecular weight (low volatility) components of the sample (like cannabinoids) in contact with the metal surface of the syringe plunger are not expelled from the syringe with the same efficiency as compounds (volatile compounds like amphetamines or solvents) whose boiling points are at or below the injection temperature. Since low injection volumes are generally used for WCOT columns (as compared with packed columns), a significant fraction of the sample is in contact with the metal surface and split discrimination may result. To prevent this, inlet liners are available to provide efficient heat transfer and thorough mixing of the sample to minimize discrimination.

Split injection is used for volatile compounds or for diluting the sample. It is also used to analyze compounds eluted rapidly after the front of solvent, and for solvent analyses after headspace preparation in order to reduce injection time. In forensic toxicology,

split injection is largely used for the analysis of medications, drugs of abuse and powders. For toxicological analyses of biological samples, splitless injection is preferred, while the split technique is most beneficial for samples containing compounds at high concentrations.

In forensic toxicology, more components are present at trace levels, and particularly in alternative matrices like hair, saliva or sweat, and target substances may be undetected because most of the sample is vented to the atmosphere and does not reach the detector. For analysis of trace compounds, the splitless injection technique is generally used.

Splitless injection may be either on-column, using a needle fine enough to enter the column bore, or off-column using a low-volume heated block. In either case, the top of the column is held at a low temperature to condense the sample contained in the solvent. Without reconcentration, the volume of the injection region will increase the band widths of eluting peaks and reduce the efficiency of the separation.

One method of reconcentration is known as the solvent effect. This occurs because the front of the solvent plug which enters the column mixes with the stationary phase and is more strongly retained than the rear of the solvent plug. Sample components therefore encounter a barrier which has the effect of condensing components at the head of the column. This applies only to compounds with a boiling point near those of the solvent. When a solvent starts to volatilize, compounds with a similar boiling point are concentrated. It is preferable to choose solvents with low volatility (i.e. hexane, toluene, octane). For example, the use of toluene for amphetamine and iso-octane for cannabinoids is recommended. Because of the interaction of the solvent and the stationary phase, some columns can be damaged if polar or aromatic solvents are used. To accomplish this solvent effect, it is necessary that the column temperature at injection is low enough to prevent the solvent from migrating too rapidly from the head of the column. This requires a column temperature of 20–40°C below the boiling point of the solvent, and may require auxillary cooling of the oven.

A second means of solute reconcentration is cold-trapping. In this method, the initial column temperature must be about 150°C below the boiling points of the components to be trapped. Compounds with lower boiling points require a solvent effect for reconcentration.

If left in the splitless condition, the time to sweep all of the solvent vapor on to the column would be sufficient to cause a large solvent tail, which would interfere with early-eluting peaks. Therefore, after a specified time (split-valve off-time), the operation of a solenoid valve causes conditions to change so that the inlet flow is greater than the column flow, which remains constant. Any solvent or sample remaining in the injector is back-flushed with carrier gas and will be purged.

On-column injection

The injection modes previously described all require flash vaporization of the sample, sometimes leading to sample discrimination or decomposition of thermally labile compounds like lormetazepam or loprazolam, two benzodiazepines. These can be overcome by injecting the sample directly on to the WCOT column through a cool injection port (at the same temperature as the column). This method of injection for WCOT columns (internal diameter 0.25–0.35 mm) was not prominent among early injector designs, owing to the mecanical difficulties of aligning the syringe needle with the column. Like splitless injection, on-column operation requires cold-trapping or the solvent effect to concentrate the sample at the head of the column. On-column injection syringes have needles which are too fine to penetrate a septum and, in order to minimize carrier gas loss, valve assemblies are used which grip the needle or which only open when the needle is in the narrow entrance channel.

Solid injection

Solid injection is used when solvent interference is serious. The ‘moving needle’ injector has found application in steroid analysis and for the determination of anticonvulsant drugs. A solution of the material to be injected is placed on the tip of the glass needle with a syringe. A small flow of carrier gas sweeps the solvent out of the top of the device to waste. The dry residue is then introduced by moving the needle into the heated injection zone of the chromatograph with a magnet. This form of injection can only be used with drugs that will not volatilize with the solvent.

Programmed temperature volatilization injection

The programmed temperature volatilization (PTV) mode of injection exhibits the advantages of split–splitless and on-column injectors. The sample is introduced in a cold vaporization chamber to avoid sample degradation and loss of compounds. The injector is warmed slowly to evaporate the solvent and to concentrate the sample (there is the possibility of concentration without the conventional extraction step with an organic phase and concentration by evaporation). The injector is then heated rapidly and split or splitless injection is operated. Degradation of analytes is minor and comparable with the on-column injection technique.

Headspace introduction

Headspace analysis permits the detection of volatile substances in a liquid or solid sample and minimizes column contamination. A small volume of the sample is placed in a vial sealed with a septum disk and this vial is equilibrated at an appropriate elevated temperature. A sample of the vapor is removed with a syringe and is then injected on to the column. This technique is used, for example, in the assay of ethanol and other solvents in blood and for complex household preparations, such as polishes, which contain volatile substances.

In forensic toxicology, biological fluids are easily and directly analyzed (without sample preparation), but tissue such as lung (identification of abused volatile substances), intestines (to determine the method of administration) and muscles may also be analyzed. Muscles are used to confirm carbon dioxide intoxication when biological fluids, and particularly blood, are not available at autopsy (for instance, if the body has been dried by the high temperature of a fire).

Methyl and ethyl mercaptans are used to document alkane intoxication (methane, propane, butane). After fatal massive ingestion of ethanol, ketones from an overloaded liver metabolism may be identified (i.e. isopropanol, acetone).

For headspace analyses, it is necessary to choose a stationary phase polarity similar to volatile polarity, for better focalization, and a large film thickness for an optimal separation of the volatile compounds.

Simple injection–double detection

Simple injection followed by double detection is possible by placing a Y quartz connector between the retention gap and two different analytical columns, each coupled to a detector. This technique can be used to analyze the large family of psychotropes (benzodiazepines, tricyclic antidepressants (TCA), neuroleptics, etc.).

Detectors

In GC, the detection problem is that of sensing a small quantity of organic compound in an inert carrier gas. To detect these compounds under such a wide variety of conditions, a number of detectors have been developed. The ideal detector will have high sensitivity, wide linear dynamic range and a small cell volume so that the GC peak is not distorted. Those in most common use are the flame ionization and electron-capture detectors.

Flame ionization detector (FID)

This detector is probably the most widely used of all the detectors because it responds to nearly all classes of compounds. Carbon compounds detectable must be capable of undergoing oxidation (hydrogen, nitrogen, water, hydrogen sulfide, sulfur dioxide, ammonia and carbon dioxide will be not detected). The principle is simple. The effluent from the column is mixed with hydrogen and the mixture burnt at a small jet in a flow of air (H_2/O_2 flame at 2000°C). Above the jet is the collector electrode. A polarizing potential (positive voltage) is applied between the jet and the electrode to measure the created ions (negative ions) when an eluted component is burnt in the flame. The FID response depends upon the numbers of ions produced by a compound.

Nitrogen–phosphorus detector (NPD) or alkali flame ionization detector (AFID)

The introduction of alkali metal vapors into the flame of a FID confers an enhanced response to compounds containing phosphorus and nitrogen. Modern versions of this detector have an electrically heated rubidium silicate source of metal ions. The detector is particularly useful for drug analysis because most drugs contain nitrogen, while the solvent and the bulk of the coextracted material from a biological sample do not. This detector is also especially useful for the detection of pesticides containing phosphorus.

Electron-capture detector (ECD)

From the time of its discovery in 1960, the ECD has enjoyed a steady growth in development and use. This is a selective detector which is highly sensitive to all electron reacting compounds containing halogen, nitro group or carbonyl group (i.e. benzodiazepines, pesticides, halogenated solvents, anesthesic gases). The detector consisted of a small chamber with two electrodes parallel to each other and a radioactive source, usually ^{63}Ni, placed close to the cathode to ionize the carrier gas. A potential applied to the electrodes produces a steady background current. When an electron-capturing substance appears in the chamber, some of the electrons are removed and a fall in the detector current results. The response of the detector is therefore a loss of signal rather than an increase in signal, as given by most other detectors.

Mass selective detector (MSD)

Generally, the courts of justice only recognize chemical–toxicological analysis results in cases where they

have been confirmed by a second independent method. The second analytical method employed is often GC–MSD. Today, GC–MSD is the method of choice as it is the more powerful tool for the identification of xenobiotics and their metabolites in specimens because of its separation ability and detection sensitivity. Mass spectrometry is based on the fact that when a molecule is ionized in a vacuum, a characteristic group of ions of different masses is formed. A mass spectrum is produced by separating these ions and recording a plot of ion abundance versus ionic mass. Mass spectrometers are classified according to the principle used to separate ionic masses. The most commonly used mass spectrometers fall into two broad classes: quadrupole and ion trap.

The most widely used method of ionization is that of electron impact (EI), in which the vaporized sample molecules are bombarded with a stream of high energy electrons. The energy absorbed causes fragmentation of the analyte, producing both negative and positive ions.

The simplest process which may occur initially is the removal of a single electron from the intact molecule to give the positively charged molecular ion (M^+). M^+ may rearrange and/or fragment into ions of lower mass-to-charge ratio (m/z), which may fragment further. The beam of ions is directed through a set of resolving slits to a detector, usually an electron multiplier. The most important piece of information that may be obtained from a mass spectrum is the molecular weight. However, certain classes of compounds do not show molecular ions.

'Soft ionization' techniques have been developed to generate a molecular ion or 'quasimolecular ion' and fragmentation is kept to a minimum. The most commonly used technique is chemical ionization (CI), in which the sample is mixed with a large excess of a reagent gas, such as methane, ammonia or isobutane. The mixture is then bombarded with high energy electrons, as in EI ionization. The reagent gas undergoes preferential ionization and the primary ions so produced react with further reagent gas molecules. These secondary ions subsequently react with the sample molecules to produce new ions. Usually, few fragmentation/rearrangement ions are observed and this quasimolecular ion is the most intense ion in the spectrum. Electronic impact and chemical ionization are complementary techniques, thus providing both molecular weight and fragmentation information. The majority of the work done has involved the study of positive ions, but interest in the negative chemical (NCI) technique has increased in recent years. For example, the CI mode of detection for negative ions is the technique of choice for detecting benzodiazepines because they possess halogen groups (electronegative functional groups) located on aromatic rings with high negative density that will give more stability to the anions formed in the ion source.

The advantage of MSD over other commonly used analytical techniques lies in the wide range of samples (blood, urine, tissues, bile, hair, saliva, gastric content, sweat, etc.) that may be examined and the amount of information that may be obtained. The main disadvantages lie in the complexity and cost of this technology.

Tandem mass spectrometry

Although GC–MS is in an advanced stage of development, new techniques, applications and instrumentations are continually being introduced. Many of these are minor modifications of existing methods, or are for specialized uses such as extended mass range and the analysis of substances of low volatility.

Among the most significant recent development is the technique of tandem mass spectrometry (MS–MS). MS–MS retains the advantages of two separations of sample components. It can be applied to a wider range of samples than GC–MS, as analytes do not have to be volatile enough to pass through a GC column. Analysis of known analytes in simple mixtures can be performed in a matter of minutes without chromatographic separation or other chemical treatment to remove interferences. To obtain increased specificity, a gas chromatograph can be added to the MS–MS configuration.

GC–MS is still the method of choice for identification of unknown compounds in a mixture. Since comparison of unknown and reference spectra is required, the many variables and configurations of MS–MS are not necessarily advantages. Spectra must be obtained under the same conditions for comparison, and large MS–MS reference files do not yet exist. For the present, MS–MS is best regarded as a specialized form of MS that is suited for certain applications not amenable to GC–MS or difficult to solve by GC–MS alone.

Pyrolysis–Gas Chromatography

There are many complex substances of forensic interest that do not have sufficient vapor pressure at the normal operating temperatures of the gas chromatograph (up to 300°C). These include hairs and fibers, paints, plastics, adhesives and other substances that are polymeric in nature. The technique of pyrolysis–GC can be conveniently used to characterize these substances.

A typical pyrolysis apparatus consists of a wand that contains a platinum coil or strip that can be

heated under controlled conditions to temperatures exceeding 1000°C. The sample can be introduced directly on to the platinum strip or inserted into a quartz tube blocked at both ends with quartz wool. The tube is then inserted into the platinum coil. The wand is then introduced into the injector of the gas chromatograph, usually via a heated interface. A controller attached to the wand controls the heating rate, duration of heating and final temperature of the pyrolysis. Since the injector and interface are continuously bathed by an inert carrier gas, the analyte does not burn (no oxygen available). Instead it decomposes into simpler molecules. If the pyrolysis is performed under constant conditions and the sample size and topography are similar each time, then the resultant pyrogram will be quite reproducible with respect to the number and relative abundance of the fragmentation. Pyrolysis–GC is also sensitive to small changes in the composition of the polymers that make up the sample, so that similar substances such as nylon 6-11 and 6-12 can be distinguished.

Applications of Gas Chromatography to Forensic Science

There are a large number of applications of GC to forensic science. The use of pyrolysis with GC has served to extend these uses.

Drugs

Drug analysis was one of the earliest uses of GC in forensic science. All sorts of 'street' drugs can be separated and quantified by GC. A few, such as amphetamines, have to be derivatized if silicon columns are employed. Blood and other body fluids are also analyzed for drugs by GC after suitable extraction processes have been carried out.

Fire residues

GC is the only major method for the analysis of fire residues. After a suitable method of concentration is applied to free the accelerants from the fire residues, GC is used to separate the components of the accelerants. In the vast majority of cases, petroleum products are used as accelerants in fires and the peak patterns from the GC analysis can be used to identify the type of product (e.g. gasoline) that was employed in the fire.

Alcohol

With the advent of computers and autosamplers, the use of GC for the detection and quantitation of alcohol in drunk-driving cases, for example, has become widespread. This is a high volume operation at many toxicology laboratories and, in many cases, the samples are loaded in the evening and run all night. If blood is used, the headspace above the blood is often analyzed, thus avoiding an extraction process.

Trace evidence

Materials such as hairs, fibers, paints, plastics and other polymers are conveniently characterized by pyrolysis–GC. This technique is very sensitive to minute differences in the chemical content of such materials, and is quite reproducible, thus aiding in comparisons of known and unknown samples.

Perspectives

Looking to the future, it is reasonable to expect continued evolutionary development in:

- sample inlets, such as automatic solid phase injectors, which do not require preparation of the specimens (hydrolysis, extraction, concentration);
- column quality, with reduction of the internal diameter in order to increased efficiency and sensitivity;
- fast separation processes using higher head-column pressure;
- detector performances – continuous improvement in sensitivity and resolving power (faster acquisition).

See also: **Analytical Techniques**: Mass Spectrometry, **Fire Investigation**: Laboratory. **Toxicology**: Overview.

Further Reading

Ardrey RE (1986) Mass spectrometry. In: Moffat AC (ed.) *Clarke's Isolation and Identification of Drugs*, 2nd edn, pp. 251–263. London: Pharmaceutical Press

Baselt RC (ed.) (1984) *Advances in Analytical Toxicology*, Vol. 1. Davis, CA: Biomedical Publications.

Gaillard Y (1998) *Methodes analytiques, Toxicology et Pharmacologie*. In: Kintz P (ed.) pp. 67–110. Amsterdam: Elsevier.

Gudzinowicz BJ and Martin HF (1977) *Fundamentals of Integrated GC/MS, Chromatographic Science Series*. New York: Marcel Dekker.

Guiochon G and Guillemin CL (1988) *Quantitative Gas Chromatography for Laboratory Analysis and On-line Process Control*. Amsterdam: Elsevier.

Jemmings W (1980) *Gas Chromatography with Glass Capillary Columns*. London: Academic Press.

Leach H and Ramsey JD (1986) *Gas Chromatography, Clarke's Isolation and Identification of Drugs*. In: A.C. Moffat (ed.) pp. 178–200. London: Pharmaceutical Press.

Maes RAA (1984) *Topics in Forensic and Analytical Toxicology*. New York: Elsevier.

McFadden WH (1973) *Techniques of Combined Gas Chromatography/Mass Spectrometry. Application in Organic Analysis*. New York: Wiley.

Porter R and O'Connor M (1974) *The Poisoned Patient. The Role of the Laboratory*. Amsterdam: Elsevier.

Sunshine I (1975) *Methodology for Analytical Toxicology*. Boca Raton, FL: CRC Press.

Supina WR (1974) *The Packed Column in Gas Chromatography, Supelco*. Philadelphia: Bellefonte.

Hyphenated Chromatographic-Spectroscopic Techniques

K P Kirkbride, Forensic Science Centre, Adelaide, South Australia

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0791

Introduction

It is possible to combine a spectroscopic technique 'on line' with a chromatographic technique to give an instrument that has a performance greater than the sum of the individual techniques. For this reason hyphenated techniques, as they are referred to, feature prominently in forensic analytical protocols.

In hyphenated techniques the specimen is subjected to chromatography in the usual manner, but the eluant is subjected to rapid and repetitive spectroscopic analysis using a dedicated instrument. The major advances in hyphenated techniques occurred when spectrometers were designed that could acquire spectral data many times per second, and when inexpensive, high-performance computers that could handle the vast amounts of data became available.

Many hyphenated techniques are available to the forensic scientist; the most important being liquid chromatography combined with mass spectrometry (MS) or ultraviolet spectroscopy, gas chromatography (GC) combined with mass spectrometry or infrared spectroscopy (IRS), and capillary electrophoresis combined with mass spectrometry or ultraviolet spectroscopy.

Although these techniques differ widely in detail, in concept their performance and utility are similar. One of the big problems with spectroscopic techniques and chromatographic techniques taken individually is a relative lack of specificity. The possibility can never be discounted that under given chromatographic conditions more than one compound elutes at a given retention time. Similarly, more than one substance might have spectral properties identical to our analyte. With hyphenated techniques the level of discrimination is raised; if more than one compound co-elutes with the unknown the chances of those compounds also having identical spectral properties is remote. Likewise if more than one compound has identical spectral properties to the unknown, then the chances of those compounds also having identical chromatographic retention is remote. For example, ethyl benzene and *m*-xylene have identical mass spectral properties, but they are easily distinguished by gas chromatography. *m*-Xylene and *p*-xylene co-elute under certain gas chromatographic conditions, but they are easily distinguished by their infrared spectral properties. Although hyphenated techniques offer very good discrimination, any one technique might not be infallible. For example, it can be difficult to resolve *m*-xylene and *p*-xylene with gas chromatography, and their mass spectral data are identical, likewise ephedrine and pseudoephedrine.

The data-handling capacity of modern computers makes hyphenated techniques very versatile. In any single chromatographic run, many thousands of spectra will be acquired (about 10 per second). These are usually all stored by the computer in the order they are acquired in an array. The data system can be interrogated in order to inspect any of the spectra acquired so that identification can be made. Such arrays are very difficult for humans to come to grips with, so the usual form of data presentation is a plot of total spectroscopic detection versus time. For example, **Fig. 1** shows a sample of the mass spectral data acquired over a short period of time during GC–MS analysis of an accelerant sample.

As each spectrum is acquired the data station stores it, and then sums the abundance of all the peaks present in the spectrum. The sum is referred to as the total ion count for that spectrum. The data station goes on to present data as plot of total ion count versus time, as shown in **Fig. 2**.

If the time axis of the plot in Fig. 2 is compressed, then the plot takes on the appearance of any other chromatogram (i.e. signal versus time, see **Fig. 3**); in GC–MS this plot is referred to as the total ion chromatogram. In a similar fashion a plot of total infrared absorbance versus time can be displayed in GC–IRS, or a plot of total UV absorbance versus time can be displayed in liquid chromatography–ultraviolet spectroscopy. It is important to understand that data are not destroyed in order to produce the total spectroscopic detection plot, it is just another form of displaying it in its totality.

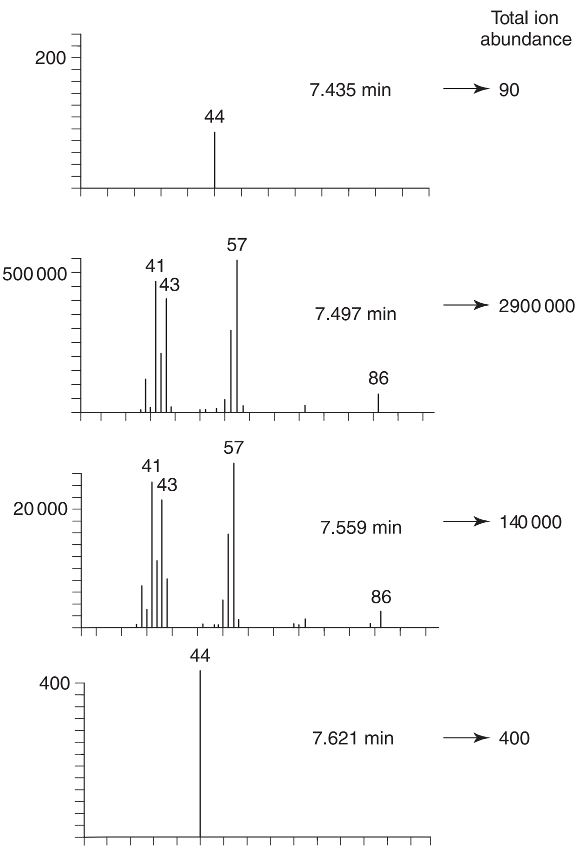


Figure 1 Pictorial representation of the construction of a total ion chromatogram from a collection of mass spectra acquired by gas chromatographic–mass spectrometric analysis of accelerant headspace. The instrument repetitively collected mass spectra about every 0.06 min. The first spectrum shown was acquired at 7.435 min and shows very low abundance of ions of mass 44, most likely due to carbon dioxide present in the carrier gas. The next spectrum acquired (7.497 min) shows many peaks; it is detecting hexane eluting from the column. The first step in the construction of the total ion chromatogram is to sum the abundances of all the peaks in each spectrum; such sums appear to the right of each of the spectra. The total ion chromatogram is then created by plotting each of those total ion abundances versus time as shown in Figs 2 and 3.

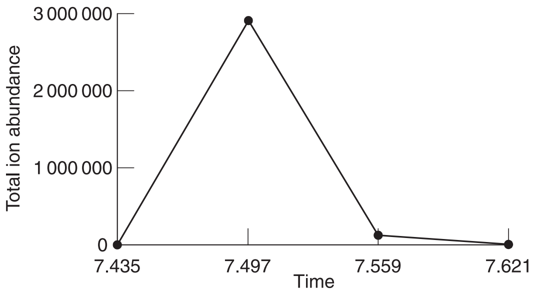


Figure 2 Total ion chromatogram produced from the data depicted in Fig. 1. Note that this chromatogram is actually a very small section of that depicted in Fig. 3.

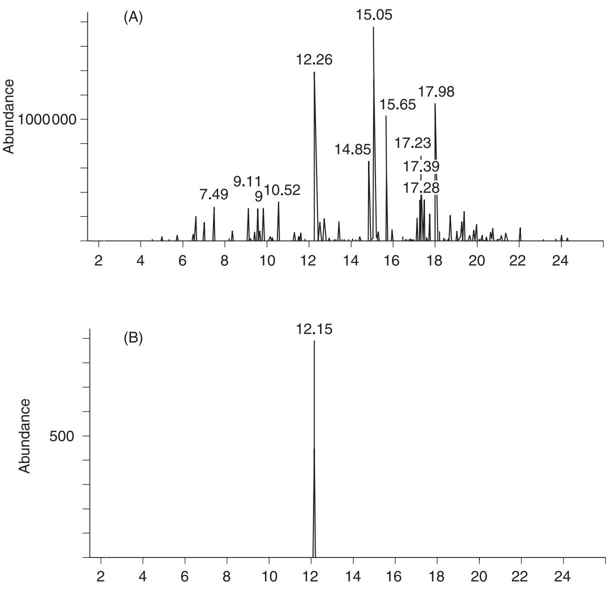


Figure 3 (**A**) Total ion chromatogram of an accelerant identified as petrol (gasoline). The chromatogram very closely resembles a gas chromatography–flame ionization detection response. Note that the chromatogram in Fig. 2 is a very small section derived from this chromatogram. (**B**) Target analysis for tetramethyl-lead using reconstructed ion chromatography. The reconstructed ion chromatogram is derived from the raw data by plotting the abundance of fragments of mass 253 versus time.

The acquired data can also be displayed in another form. Consider the example of GC–MS as applied to accelerant identification. Alkylated aromatic compounds (i.e. toluene, ethyl benzene, xylenes, propyl benzene etc.) always fragment under mass spectrometry to yield a strong peak at 91. It is possible to instruct the data station to sort through all spectra acquired during a chromatographic run and plot only the abundance of 91 versus time. What is produced is referred to as a reconstructed ion chromatogram (or an extracted ion chromatogram, or a mass fragmentogram). The reconstructed ion chromatogram resembles a standard gas chromatogram, but in this instance it shows only those peaks arising from aromatic compounds. Reconstructed ion chromatography can be applied in any situation where a class of compounds exhibit a similarity in fragmentation pattern. It is a very useful technique for identifying a pattern in a complex collection of chromatographic peaks, such as the chromatogram of an accelerant contaminated with pyrolysis products. This concept can be applied to other hyphenated techniques, for example complicated gas chromatography–infrared data can be reduced to indicate the presence of certain classes of compound.

Reconstructed ion chromatography can also be very useful for target analysis if the target has some unique mass spectral characteristic. A good example is

the detection of lead antiknock compounds in accelerants. Tetramethyl-lead gives a strong peak at 253 in its mass spectrum. Fig. 3 shows a total ion chromatogram for an accelerant identified as petrol (gasoline). Fig. 3 also shows a reconstructed ion chromatogram derived from the accelerant where the data station is instructed to plot abundance of 253 versus time. As can be seen, a peak near 12 minutes is obvious in the reconstructed ion chromatogram, but it is not obvious in the total ion chromatogram. Reconstructed ion chromatography therefore highlights the presence of target compounds that otherwise might be overlooked in a total ion chromatogram. Another good example is in the detection of explosives residues. Total ion chromatograms derived from bombing debris extracts often contain many irrelevant peaks. Nitroesters (i.e. nitroglycerin, ethylene glycol dinitrate, pentaerythritol tetranitrate) all give a strong peak at 46. Peaks in a reconstructed ion chromatogram (abundance of mass 46 versus time) indicate whether any of those explosives are present in the extract. Further tests using explosives standards can then be undertaken in order to confirm peaks on the basis of their retention time. Likewise, if a compound has a unique infrared peak, it can be highlighted in target analysis using gas chromatography–infrared spectroscopy.

Reconstructed ion chromatography, or its equivalent in other hyphenated techniques, can be used to 'resolve' compounds that overlap or even totally co-elute during chromatography if the compounds have some unique spectral properties. For example, under some gas chromatographic conditions heroin and phthalate ester plasticizer co-elute; however, the plasticizer has a strong peak at 149 in its mass spectrum whereas heroin does not, likewise heroin has a strong peak at 369 but the plasticizer does not. A reconstructed ion chromatogram using 149 will show the peak due to plasticizer, whereas the chromatogram using 369 will show the peak due to heroin, thus the peaks can be 'resolved' even though they co-elute.

This reasoning can be extended further using GC–MS to create a target analytical technique with very low limits of detection called selective ion monitoring. In this technique one must have a target compound in mind, and its mass spectral data must be known accurately. During chromatography the mass spectrometer is instructed to dwell on one or only a few selected mass peaks, for example 369 in the case of heroin. As the mass spectrometer is concentrating only on a very restricted range it can be tuned to be very sensitive to them, therefore limits of detection can be very low. In this case if the matrix does not contain any compounds that yield peaks at 369 then the technique will display only heroin, even if that drug co-elutes with many other compounds. It is even possible to perform quantitative analysis. Selective ion monitoring therefore finds great use in trace target analysis where analytes are well characterized by mass spectrometry, such as detection of explosives in dirty debris, or detection of low dose drugs in human fluids.

See also: **Analytical Techniques**: Mass Spectrometry; Spectroscopic Techniques. **Explosives**: Analysis.

Further Reading

Cole RB (1997) *Electrospray Ionization Mass Spectrometry: Fundamental Instrumentation and Applications*. New York: Wiley.

Karasek FW and Clement RE (1988) *Basic Gas Chromatography–Mass Spectrometry: Principles and Techniques*. Amsterdam: Elsevier.

Message G (1984) *Practical Aspects of Gas Chromatography/Mass Spectrometry*. New York: Wiley.

Yinon J and Zitrin S (1993) *Modern Methods and Applications in Analysis of Explosives*. Chichester: Wiley.

Mass Spectrometry

J Schuberth, Lidingö, Sweden; formerly Head, National Laboratory of Forensic Chemistry, Linköping, Sweden

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0413

Introduction

To reveal a crime the forensic scientist looks for marks a felon may have left at the site of the event or brought with him. These are often comprised of agents suitable for chemical analysis, e.g. drugs seized by the police or present in a body fluid, residues of arson accelerants or of explosives, poisons in a dead body to name a few examples. A main problem is that the chemist does not know which compounds to look for in a wide range of possible ones with different physicochemical properties. He has to search for the 'general unknown', a circumstance that requires not only solid professional skill but also good test tools. One of the chemist's main analytical aids to solve forensic problems by revealing the 'general unknown' is the use of mass spectrometry (MS).

On the basis of experiments on the behavior of charged particles in electrical and magnetic fields, J. J.

Thomson built the first mass spectrometer known as the hyperbola spectrograph. Since its advent in 1913 several types of mass spectrometer have been used as the most versatile apparatus in diverse analytical situations to detect and identify virtually any analyte molecules that can be converted to the gas phase at a reasonably low pressure. Also, the method provides an excellent means of finding out the molecular weights and gaining structural information on analytes, even of those larger than 10 000 Da.

In short, MS is a method by which one generates charged molecules and molecular fragments, and measures each of their mass, or more correctly, the mass/charge ratio. The uncharged gas molecules, which are introduced into the mass spectrometer, move around randomly in space, and to control their motion they have to be ionized, i.e. a charge is added to the molecules. To obtain structural information about the molecules, energy in excess of the ionization is supplied to break some of the covalent bonds holding the molecules together.

Even though mass spectrometers work differently to reach these goals, they have some general operating concepts in common. These are shown in **Table 1** and consist of the inlet system, ion source, mass analyzer, mass detector, data collecting system, and data interpretation scheme.

Inlet System

The inlet system is used to introduce the sample into the mass spectrometer, to convert it into the gas phase, and to reduce its pressure before ionization. Forensic samples are often impure, and the analytes, therefore, have to be separated from the matrix before insertion into the mass spectrometer. This is done by chromatographic methods, i.e. thin layer chromatography, gas chromatography (GC), or high-pressure liquid chromatography (LC). After thin layer chromatography the cleaned-up sample is introduced into the mass spectrometer by the direct insertion probe. The specimen to be analyzed is put on a metal rod, which is inserted into the mass spectrometer and then heated to evaporate the sample close to the ion source. Provided that the contaminants in the forensic samples have boiling points which differ sufficiently from that of the analyte, some separation may yet be possible by slowly increasing the temperature of the rod. The method is rather imprecise making it inappropriate for quantitative study; the technique of direct inlet insertion is therefore seldom used for MS analysis of forensic samples.

More suited is the use of the inlet system as an interface between a chromatographic device and the mass spectrometer. By this technique the analytes are separated from one another and from the contaminants by either GC or LC, and the effluents from the column flow directly into the mass spectrometer. The on-line coupling of a gas chromatograph to the mass spectrometer was the first interface to be developed. It functions in a quite different way than the LC-MS interface, which for forensic aims is less used than the GC-MS interface.

The LC-MS interface has a number of problems. The main reason for this is the fact that the eluate from the column is a liquid, and on changing to a gas for MS analysis it expands, thus creating an extra burden on the vacuum system for reducing the pressure in the instrument. Moreover, the effluents often carry polar, heat-labile substances that may taint the ion source when they are vaporized.

With the most often used inlet of the LC effluents, i.e. the thermospray interface, these are forced through a pinhole leak to form a jet of liquids, which is heated to vaporize the solvents in the aerosol. The mist of droplets in the gas phase, which becomes void of solvents, carries along with a charge from any ions initially present in the solution the less volatile analytes directly into the ion source of the mass spectrometer to become fit for analysis.

GC-MS was devised about 25 years ago, and in those days the GC separation was done with packed columns. Since small amounts of the analyte are then eluted from the column in a large volume of carrier

Table 1 Operating concepts of mass spectrometers

Inlet system	*Ion source*	*Mass analyser*	*Detector*	*Data collecting system*	*Data interpretation*
Direct inlet	Electron impact	Magnetic sector instrument	Electron multiplier	Full scan mode	Library searching
Gas chromatography	Chemical ionization	Quadrupole instrument	(Time-of flight)[a]	Selected ion monitoring	(Artificial intelligence)
Liquid chromatography	(Desorptive ionization)	Ion-trap detector Tandem MS		Negative ion detection	

[a] The techniques in parentheses are not commonly used in forensic science, and are not, therefore, addressed in the text.

gas at atmospheric pressure, whereas the mass spectrometer works at about 10^{-3} Pa, the interface has both to lower the pressure of the sample before it enters the mass spectrometer, and also to concentrate the analyte. This is achieved with a molecular separator, in which the effluent from the gas chromatograph, made up by the carrier gas of lower mass than that of the analyte, is pressed through a capillary with a tight outlet into a nearby capillary with a wider orifice. By this procedure the lighter molecules are more readily lost into the vacuum area, whereas the heavier analyte molecules move through the second capillary into the mass spectrometer, which they reach after having been enriched.

Today in most gas chromatographs, however, the packed column is replaced with a capillary column to do the separation. This has to a great extent made the problem of introducing the analytes into the mass spectrometer easier. Since the flow rates are much lower in a capillary than in a packed column, the whole volume of the effluents can be let into the mass spectrometer without any losses in the interface. To do this, the end of the GC capillary is close to the ion source of the mass spectrometer. With this system there are two things the user must keep in mind. One is that with a magnetic sector or a quadrupole instrument a vacuum pump must be used with enough capacity to guard an adequate pressure in the ion source. With an ion-trap, which also needs a vacuum pump, this aim is reached by fine tuning the flow rate through the capillary. The other factor is to keep the interface at a raised temperature to avoid analyte condensation in the liner.

Ion Source

When the analytes enter the mass spectrometer's ion source these are submitted to ionization. The main aim of this step is to put a charge on the molecule, a process needed to enable a control of the molecule path in a magnetic or oscillating electric field. Another goal is to break some of the chemical bonds of the molecule by putting energy into it.

The simplest ionization method is to impinge the analyte molecule (M) with a beam of energetic electrons (e^-), a process referred to as electron impact (EI), and which results in the formation of a radical cation with an odd number of ions ($M^+.$) according to the following equation.

$$M + e^- \rightarrow M^+. + 2\,e^-$$

As seen the analyte molecule has lost an electron and become a particle with a positive charge, thus allowing it to be separated in a magnetic or oscillating electric field based on its mass and number of charges. The energy of the electrons is generally set at 70 eV, an optional value chosen because it is high enough to exceed the ionization energy at about 10 eV, and to split the molecule by breaking its chemical bonds.

For the identification of an analyte it is often of utmost importance to know its molecular weight. However, this information is generally not gained by ionization with electron bombardment; the process often results in a complex fragmentation of the molecule, making the deduction of the original substance mass difficult. To overcome this problem a milder ionization method has been devised, i.e. chemical ionization (CI). In this process the analyte molecule reacts in a gas phase with a proton donor formed by EI of a reagent gas, e.g. methane. The reactions occur according to the following.

$$CH_4 + e^- \rightarrow CH_4^+. + 2e^- \text{(primary ion)}$$
$$CH_4^+. + CH_4 \rightarrow CH_5^+ + CH_3^{\cdot} \text{(secondary ion)}$$
$$CH_5^+ + M \rightarrow CH_4 + [M + 1]^+$$

By accepting a proton and taking up a charge, the analyte molecule, thus, reacts with a secondary ion formed by EI of the reaction gas, methane. The process now allows the analyte molecule to become available for MS analysis and to show up in a mass spectrum with its protonated variant as the most abundant peak.

Mass Analyser

The analyte molecules, which have become ionized, may now be moved apart from one another in a magnetic or oscillating electric field. Even though the main object is to analyse the fragments based on their masses, it is the mass/charge (m/z) ratio that forms the ground for the separation process. This means that a fragment with m/z = 200/2 will have the same trajectory as a fragment with m/z = 100/1, and they may, thus, not be distinguished from one another. Fragments with two charges are, however, rare, and in practice it is the mass of a fragment that one generally regards as the controlling factor of its movement in a magnetic or oscillating electric field.

Magnetic sector instrument

The first commercial mass spectrometer was a magnetic sector instrument. Charged fragments or molecules, which are accelerated and ousted from the ion source, tend to adjust to orbital movements when allowed to travel in a magnetic field aimed perpendicular to the flight of the ions. The trajectory of a fragment with a given m/z value, thus, depends on the acceleration voltage and the strength of the magnetic field. This means that only those fragments, which have a m/z value to balance the centripetal force of

the magnetic field and become stable at a given acceleration, will travel through the flight tube and reach the detector, whereas the unstable ions will adhere to the path wall.

The fragments exit the flight tube through a slit, whose width determines the analytical resolution and the sensitivity. A wide opening favors a high sensitivity but low resolution. A small aperture, on the other hand, gives a high resolution but low sensitivity, and a reason for this is a certain spread of fragments with the same *m/z* value but with different kinetic energies. These fragments will, thus, not be recorded. To compensate for such a loss one may insert either before or after the magnetic sectors, a device to focus the fragments with the same *m/z* value, i.e. an electrostatic analyzer. This procedure forms the basis for the double focusing mass spectrometer.

Quadrupole instrument

Another type of mass filter is the quadrupole. Its ability to separate fragments with different masses is based on the fact that, when ions travel in a beam of an oscillating electric field, their trajectories become influenced. The quadrupole mass filter is made up of four parallel rods, about 10 cm long and 1 cm in diameter, which in cross-section are arranged to form a square box with a gap in between the four bars. The fragments are allowed to travel in this space along the rods, where they are accelerated in the oscillating electric field set up by dc and ac potentials applied to the rods. At a certain ac and dc potential, ions with a specific *m/z* value will become stable and oscillate in a fixed path through the quadrupole to reach the detector to be recorded. The fragments with *m/z* values that do not suit the applied potentials will become unstable, and, as in the magnetic sector instrument, not reach the detector but stick to the rods.

Ion-trap detector

A third type of mass filter is the ion-trap detector (ITD). Like the quadrupole, the ITD operates on the concept that when ions travel in an oscillating electric field their trajectories become influenced by the wave frequency. Unlike the quadrupole, however, the ITD uses a closed cavity in which the electric field is set up by the dc and ac potentials applied to a ring electrode and an end cap of the cell. The molecules enter the cell where they are ionized by an electron beam, and under the control of the applied electric field the ions are forced to move in an orbit within the space of the cell. When the ac or dc potential is changed, the motion of some ions becomes unstable, and 50% of these are then ousted from the cell through a hole in its bottom to be recorded by a detector. Unlike the quadrupole or magnetic sector instrument in which the stable ions are registered, the ITD, thus, looks at the unstable ions. This means, that compared with the quadrupole or magnetic sector system, the ITD offers a rather different analytical concept.

Tandem mass spectrometer

To gain a higher selectivity and also more information about the nature of an analyte, its analysis can be done by tandem mass spectrometry (MS/MS). Another main advantage of MS/MS is its ability to reduce the chemical background noise, and, thus to increase the signal-to-noise ratio for a detected peak. Although this method, as judged from the number of scientific papers reported, does not seem to be in common use for forensic analysis today, MS/MS will probably become the method of choice in the future. It may offer the forensic chemist a powerful tool for the analysis of the complex mixtures presented. In MS/MS two mass spectrometers are linked in series to each other, in the early apparatus set up by two magnetic sector instruments. However, since these instruments are very bulky and high-priced, modern tandem devices are nearly all based on the use of quadrupoles or ITDs.

A dual quadrupole mass spectrometer actually entails three quadrupole units. In the first unit the incoming parent ions are separated from one another. In the second unit, which is charged only with radio frequency but not with dc the parent ions are taken to a common focus without any extra mass separation, and, similar to the mode of chemical ionization, split into daughter ions by collision with a gas let in from the outside. In the third unit the daughter ions are mass filtered as described above under 'Quadrupole instrument', and a fragmentation pattern in addition to that of the parent ion can be formed from each one of these.

A rather recent MS/MS development is the use of the ITD. To call the device for this approach a tandem mass spectrometer may, however, be somewhat deceptive. Even though the MS/MS analysis is done in three steps as described above for the quadrupole MS/MS, these are all carried out at the same site, i.e. within a single ITD cell. In the first step the parent ions, which are to be submitted to further analysis, are stored in the cell. Unlike the conventional ITD mode, when the target ions to be analyzed are the unstable ones, the MS/MS operation now saves the stable ones, whereas the unstable ions are discarded from the cell. In the second step the selected parent ions, moving in an orbital, are allowed to collide with a gas let into the ITD cell, whereby the daughter ions are formed. While traveling in the oscillating electric field, some

of these become unstable at a specific change in the ac or dc potential applied on the ITD cell, and are then finally ejected to the detector to generate a daughter-ion spectrum. Since this MS/MS approach is carried out in a single compartment, rather than in three separate ones, there is a saving in cost.

Detector

In most instances, the separated ions are detected with an electron multiplier. Its task is to convert the charged fragments into electrons, amplify these, and transfer the electrical current set up to a wire for recording as a signal. Usually a so-called horn-type electron multiplier is used, since this variety of detector is the most compact and low-cost device. When the charged fragments enter the detector and strike the surface area of the horn, electrons are emitted. After acceleration of the electrons by an electrical potential difference in the horn, they in turn hit the surface and new electrons are formed, a process that is repeated over and over again to generate a cascade of progressively raised number of electrons. Usually the gain of emitted electrons is in the order of 10^4–10^7 per ion entering the detector.

Data Collecting System

The MS test can be run in the full scan mode or in the selected ion monitoring (SIM) mode. In the full scan mode the mass analysis covers a range of *m/z* values, whereas in SIM a limited number of *m/z* values are selected for the experiment. When choosing between the two options, the operator of the mass spectrometer has to decide to what extent he is willing to trade sensitivity for selectivity or vice versa. This consideration is based on the kind of study, i.e. a search for the ‘general unknown’ or for the suspected agent. If no specific substance is expected to be present in the sample, he may be bound to screen for a number of candidates, e.g. for drug substances in a toxicology survey or for accelerants in an arson study. In such a situation the mass spectrometer is run in the scan mode, which means that the apparatus is set for recording a number of *m/z* values in a wide mass range. By this approach a number of unforeseen agents can be detected, but, unfortunately, at the sacrifice of sensitivity. On the other hand, if a specific compound is to be verified, the mass spectrometer can be set to focus on a single *m/z* value or a limited number of ions that are formed from the expected find in the sample. This is an analytical approach that aids a high sensitivity, but also yields only a narrow range of substances.

The analytical condition the operator has to consider also depends on the type of mass spectrometer being used, i.e. the quadrupole, the magnetic sector instrument or the ITD. From the analyst’s point of view, the quadrupole and the magnetic sector instrument, which are so-called beam-type scanning mass spectrometers, offer rather small differences. They both operate on the principle of mass-selective stability. This means that ions with a small range of *m/z* values, conforming with the magnetic instrument’s given magnetic field strength and acceleration voltage, or with a quadrupole filter’s oscillating electric field, will be transmitted through the discriminating device into the detector. By this means the ions with stable trajectories become detectable, whereas those with the unstable ones are trapped on the magnet or on the quadrupole rods. A vital aspect of this operation mode is that the ionization and the detection occur, as a continuous process not separated in time, a circumstance that to a large extent affects the sensitivity. The reason for this is that the window of the stable *m/z* values is sequentially swept across the entire *m/z* range of interest, and the rate of the window width/entire *m/z* range, i.e. the duty cycle, is in most scanning tests only a fraction of 1%. More than 99% of the ions from the target agents are, thus, lost. A duty cycle of 100% is yet possible with a beam-type instrument but only when run in a nonscanning mode, as in the SIM. Such an application, though, requires that the analyst know what he is looking for in the sample.

As described earlier the ITD monitors the ions with unstable trajectories, and, thus, operates according to the mass-selective instability notion. This means that, contrary to the beam-type device, the ITD works in two steps, i.e. ion accumulation and mass analysis. Since these are separated in time, the yield of detectable ions will become high and rather independent on the scan range. The main advantage of the ITD over the beam-type scanning approach is that the former may give nearly the same sensitivity when run in the scan as in the SIM mode. The trade-off between the extent of the mass range chosen and the test sensitivity gained, thus, becomes less crucial with the ITD than with the magnetic or quadrupole device. By the use of the ITD the analyst may, thus, scan with a high sensitivity the whole mass range that covers a substance group of interest. Given that the forensic scientist often does not know what to look for, and, therefore, needs a search method with a high sensitivity, the ITD should perhaps best meet his demand. A drawback of the ITD is that its sensitivity is more dependent on interfering substances than the beam-type scanning mass spectrometer; the ITD sensitivity, thus, drops with increasing amounts of impurities that may be co-eluted with the analytes during

the chromatographic separation. The generation of somewhat distorted mass spectra at high analyte concentrations giving rise to enhanced $[M+1]^+$ peaks is another inherited ITD problem. The reason for this may be an overload of the ion source, an event that probably gives rise to some chemical ionization by the analyte molecules themselves.

In addition to the positive ions formed during ionization of a molecule by EI or CI, negative ions are also produced, and by changing the electric field of the mass spectrometer these can be monitored. In certain instances, i.e. when the target substances have a high affinity for electrons, negative ion monitoring can be extremely useful, mainly because of the high sensitivity that can be achieved. However, such a situation is only reached when CI, often resulting in a low fragmentation with the formation of an ionized molecule, keeps the energy of the electrons at a low level. At high electron energy, on the other hand, a number of low-mass fragments without much analytical information are formed. The approach using negative ion monitoring has been particularly fruitful for the analysis of halogenated drug substances, which have been detected at 100–1000-fold higher sensitivity than when tested by positive ion monitoring.

Data Interpretation

A pencil and paper are the classical tools the analyst uses to decode a mass spectrum for pin pointing the agent, which may be the origin of a recorded fragment pattern. By this means he tries to postulate the substance based on the features of the spectrum along with his knowledge about basic reaction mechanisms. The approach, however, entails a number of steps to reveal the elemental make-up for the main ions, the molecular ion, and the functional groups from series of ions or characteristic ions showing up in the spectrum. These efforts are done along with trial and error to predict the major fragments from often a variety of postulated structures. It is obvious that this is a clumsy and time-consuming method, which calls for great skill from the analyst, and also allows some decipherable bias. It is, thus, not suited for the everyday work.

Today the evaluation of a recorded mass spectrum is computerized. By comparing the mass spectrum revealed with reference mass spectra in a database, one looks for the agent with a spectrum that best matches that of the unknown analyte. For collating, different algorithms have been devised, and in most instances the data program produces ten candidates. Each of these is rated with a number between 0.0 for a very poor fit and 10.0 for a perfect match to the mass spectrum of the analyte. Even though the search system is based on the notion that similar mass spectra of the reference and of the unknown substance mark a chemical unity, the use of such a method requires that the operator look upon the search data with a great deal of skepticism. A common problem is that there is seldom a perfect match between the mass spectra of the analyte and of the reference agent in the library. This may be because the unknown substance is present only in low concentrations and some of its fragments, therefore, do not show up, or it is tainted by impurities, which make the recorded mass spectrum unsure. Also, the library may lack a corresponding reference substance or yield only abridged mass spectra to reduce the program's search time among the often several tens of thousands of entrainees. These are situations that will make a perfect fit unlikely. Even though a computerized interpretation of mass spectra offers the operator a great help, a final verification must be made by some independent means. Preferably, a mass spectrum of the reference substance should be run under the same conditions as used for the sample test, alternatively the retention time or the retention index value of the analyte should be compared with that of the candidate compound arrived at after the mass spectrum search.

Some Forensic Applications of MS

The search for drug substances and poisons in body fluids from living persons or in organs of postmortem materials offers an important and difficult task for the chemist. In a forensic survey toxicology perhaps yields the most common use of MS, and its utility in such work is elaborated on elsewhere. It shows an example of GC/MS screening of a urine sample for drug substances with the apparatus run in the scan mode and with the library search for compounds of the main peak showing up on the mass chromatogram. As pointed out earlier, only rather high concentrations of the 'general unknown' can be spotted under such test conditions.

In many instances the search of a forensic sample for trace amounts of the 'general unknown' requires that a MS system with higher sensitivity be used than offered by the mass spectrometer run in the scan mode. The analyst of a fire debris sample is often faced with such a need. The aim of his work is to detect and identify in a fire debris sample residues of accelerants that could have been used by an arsonist to start the fire. Since the substances screened for are volatile and, thus, can be refined by headspace extraction into a rather clean blend, the use of an ITD is seemly. **Fig. 1** shows the steps involved in the MS analysis of a fire sample from a real-life event.

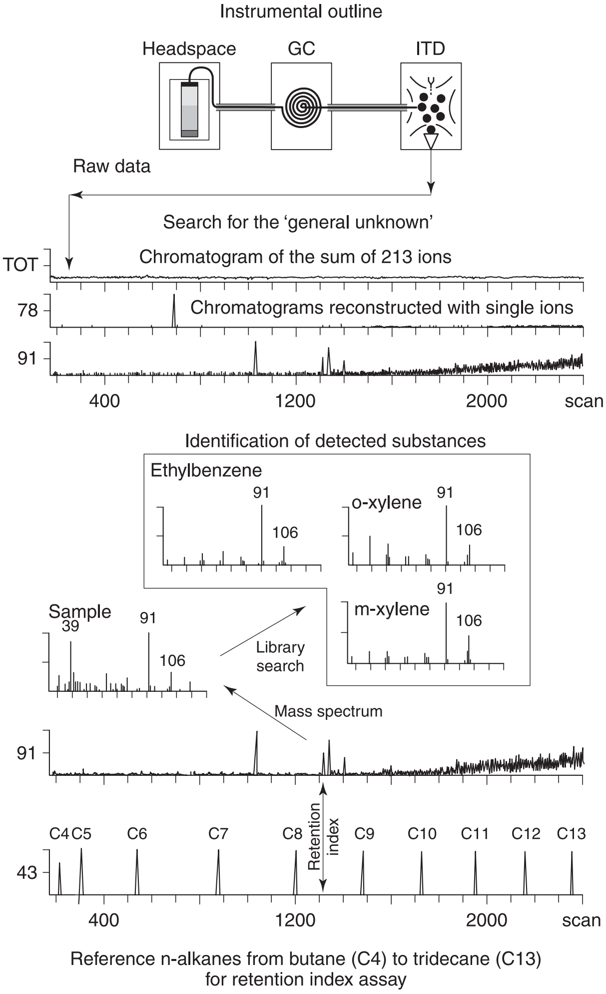


Figure 1 GC/MS search for arson accelerents in postmortem sample of actual fire victim. As shown under 'Instrument outline' volatile organic compounds released during the headspace extraction of the sample are chromatographed by GC, and the isolated components let directly into the ion source of the ITD. In the 'Search for the general unknown', the sum of all 213 ions registered is monitored as the total ion current (TOT). No peaks can be seen at this point. In the next step 213 new chromatograms are reconstructed with each single ion, two of which are shown in the figure at m/z 78 or 91. Since a more favorable signal-to-noise ratio is obtained, a number of peaks with yet unknown identities can be seen on both reconstructed mass chromatograms. As shown under 'Identification of detected substances', the second peak on the mass chromatogram reconstructed with ion at m/z 91 is taken as an example. Its mass spectrum is recorded and compared with the mass spectra in a library. The mass spectrum quality of the unknown compound is, however, not good enough for picking a clear match among the three 'hottest' library candidates shown in the figure. To make a final identification the mass spectrum data are supplemented with retention index values indicating that the unknown substance is ethylbenzene.

See also: **Drugs of Abuse**: Analysis. **Fire Investigation**: Types of Fire; Physics/Thermodynamics; Chemistry of Fire; Evidence Recovery at the Fire-scene; Fire-scene; Fire-scene Patterns. **Toxicology**: Overview. **Analytical Techniques**: Hyphenated Chromatographic-Spectroscopic Techniques.

Further Reading

Chapman JR (1993) *Practical Organic Mass Spectrometry: a Guide for Chemical and Biochemical Analysis*, 2nd edn. Chichester: Wiley.

Constantin E and Schnell A (eds) (1990) *Mass Spectrometry*. New York: Horwood.

Harrison AG (1992) *Chemical Ionization Mass Spectrometry*, 2nd edn. Boca Raton: CRC Press.

Horna A (1993) Chemical ionization mass spectrometry and its utilization for the analysis of organic substances. In: Churácek J (ed.) *Advanced Instrumental Methods of Chemical Analysis*, pp. 352–378. New York: Ellis Horwood.

Karasek FW and Clement RE (eds) (1988) *Basic Gas Chromatography–Mass Spectrometry*: *Principles and Techniques*. Amsterdam: Elsevier.

Masucci JA and Caldwell GW (1995) Techniques for gas chromatography/mass spectrometry. In: Grob RL (ed.) *Modern Practice of Gas Chromatography*, 3rd edn. New York: Wiley.

McLafferty FW and Turacek F (eds) (1993) *Interpretation of Mass Spectra*, 4th edn. Mill Valley: University Science.

McNair HM and Miller JM (1998) Special topics. In: *Basic Gas Chromatography. Techniques in Analytical Chemistry*, pp. 153–163. New York: Wiley.

Rule G, McLaughlin LG and Henion J (1993) A quest for oleandrin in decayed human tissue. *Analytical Chemistry* **65**:857A–863A.

Russel DH (ed.) (1994) Experimental mass spectrometry. In *Topics in Mass Spectrometry*, vol. 1. New York: Plenum.

Scott RPW (1996) Chromatographic detectors. Design, function, and operation. In: *Chromatographic Science Series*, vol. 73, pp. 380–388. New York: Marcel Dekker.

Silverstein RM, Bassler GC and Morrill TC (1991) Mass spectrometry. In: *Spectrometric Identification of Organic Compounds*, 5th edn. pp. 3–89. New York: Wiley.

Smith RM (1982) Arson analysis by mass chromatography. *Analytical Chemistry* **54**:1399–1409.

Microscopy

S Palenik, Microtrace , Elgin, IL 60123, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0410

Introduction

The applications of microscopy in the forensic sciences are almost limitless. This is due in large measure to the ability of microscopes to detect, resolve and image the smallest items of evidence, often without alteration or destruction. As a result,

microscopes have become nearly indispensable in all forensic disciplines involving the natural sciences. Thus, a firearms examiner comparing a bullet, a trace evidence specialist identifying and comparing fibers, hairs, soils or dust, a document examiner studying ink line crossings or paper fibers, and a serologist scrutinizing a bloodstain, all rely on microscopes, in spite of the fact that each may use them in different ways and for different purposes.

The principal purpose of any microscope is to form an enlarged image of a small object. As the image is more greatly magnified, the concern then becomes resolution; the ability to see increasingly fine details as the magnification is increased. For most observers, the ability to see fine details of an item of evidence at a convenient magnification, is sufficient. For many items, such as ink lines, bloodstains or bullets, no treatment is required and the evidence may typically be studied directly under the appropriate microscope without any form of sample preparation. For other types of evidence, particularly traces of particulate matter, sample preparation before the microscopical examination begins is often essential. Typical examples of sample preparation might include: mounting a crystal in index of refraction oils to determine its optical properties, reconstituting a blood crust particle and staining it for leukocytes and other cells, preparing a cross-section of a fiber or mounting a specimen in a nonfluorescent mounting medium to observe its autofluorescence. As a general rule, the type of specimen, the information one wishes to obtain from it and the type of microscope chosen for the task will determine if sample preparation is required and the type of processing required.

Types of Microscopes Used in the Forensic Sciences

A variety of microscopes are used in any modern forensic science laboratory. Most of these are light microscopes which use photons to form images, but electron microscopes, particularly the scanning electron microscope (SEM), are finding applications in larger, full service laboratories because of their wide range of magnification, high resolving power and ability to perform elemental analyses when equipped with an energy or wavelength dispersive X-ray spectrometer.

Stereomicroscope

This is the simplest type of microscope in terms of both construction and use. The stereomicroscope consists of two compound microscopes which are aligned side-by-side at the correct visual angle to provide a true stereoscopic image. The long working distance (space between the specimen and objective lens), upright nonreversed image and large field of view make these the instruments of choice for performing preliminary examinations of evidence as well as manipulating small particles and fibers to prepare them for more detailed microscopical or instrumental analyses or comparisons. An additional advantage which results from the long working distance and illumination by reflected light is that specimens rarely require any sample preparation. The specimen is simply placed under the microscope and observed.

The useful magnification range of stereomicroscopes is typically between $2.5\times$ and about $100\times$. Modern stereomicroscopes incorporate a number of features which increase their utility and ease of use. A choice of illuminators which can provide brightfield and darkfield reflected, fluorescence and transmitted light permit the microscopist to visualize microscopic objects and features which might otherwise appear invisible, and thus escape detection. Attaching the microscope to a boom stand permits it to be swung out over large objects such as clothing, piles of debris or even entire vehicles. Both photographic and video cameras can be attached to record images for inclusion in a report, as a courtroom exhibit or to display to colleagues. Even the least experienced members of the laboratory staff can use these instruments with very little training.

Compound microscope

Compound microscopes represent a significant step up in magnification, resolution and difficulty of use from the stereomicroscope. Magnifications range from $2.5\times$ to about $1300\times$ with a corresponding increase in resolving power. Most observations with these instruments in the forensic science laboratory are made with transmitted light which places limitations on the specimens which are to be studied. Reflected light instruments, with the exception of fluorescence microscopes and comparison microscopes used to study bullets and tool marks, have found limited use in forensic laboratories and are generally confined to the examination of metals that have been prepared by grinding and polishing. For transparent specimens, sample preparation becomes significant, not only because specimens must be thin enough to transmit light, but also because these methods may introduce artifacts that must be recognized when performing both identifications and comparisons. A variety of compound microscopes are available to the forensic microscopist and their selection will depend on the types of evidence to be studied. These include standard brightfield, phase

contrast, comparison, hot stage, fluorescence and polarizing microscopes.

Brightfield microscopy is used to observe and study the morphology of microscopic specimens. In the forensic laboratory these can include a range of materials almost too numerous to list. These are substances which must be identified on the basis of their microscopic morphology or, having been identified by other means, exhibit microscopic features which can aid in their comparison to a suspected source. Several examples are listed along with typical sample preparation methods in **Table 1**. Sample preparation for transmitted light brightfield microscopy may be as simple as mounting the specimen in a temporary or permanent mounting medium to achieve a particular level of contrast. Staining may also be used to provide contrast or demonstrate the presence of particular chemical compounds or functional groups. It may also involve physical manipulation such as preparing a replica or cutting a cross-section. The selection of the proper method of preparation for a particular specimen will depend on the nature of the sample itself and the information which the microscopist is trying to obtain from it. The ability to choose and carry out the best methods of sample preparation, which are appropriate to both the small size and irreplaceable nature of these specimens, is one of the hallmarks of the expert forensic microscopist.

The *polarizing microscope* is arguably the most useful and versatile instrument in the hands of a trained and experienced forensic microscopist. Not only does it perform all the duties of a normal brightfield microscope for the study of morphology, but it also permits observations and measurements in plane polarized light and between crossed polars. Polarized light microscopy provides both qualitative and quantitative information which is of value in observing, identifying and comparing microscopic particles, crystals and fibers. The principal components which distinguish a polarizing microscope from a conventional brightfield instrument are two polarizers which are inserted above and below the specimen, a circular rotating stage graduated in degrees, and a slot for the insertion of compensators. Compensators are used to introduce interference colors into specimen images for contrast and to determine certain qualitative and quantitative optical properties of crystalline solids. The polarizing microscope allows transparent solids to be examined in plane polarized light (to isolate unique vibration directions in a crystal or crystalline polymer), between crossed polars (to observe and measure birefringence and to

Table 1 Examples of the identification of various types of evidence by microscopic morphology

Type of evidence	*Sample preparation required*	*Identification features observed*
Hairs	Preparation on glass slides in mounting medium. Scale impressions. Cross-sections. Clearing medulla.	Color. Overall shape and length. Scale patterns. Cross-sectional shape and structural features. Pigmentation distribution. Cellular structure of medulla.
Vegetable fibers	Scraping to look for epidermal and vascular tissue and crystals. Maceration into fiber ultimates. Microchemical testing for lignin content.	General color and texture. Shape of any crystals recovered. Appearance of epidermis and other cells. Features of fiber ultimates: length, structure of lumen, cell walls and pits. Comparison to reference slides.
Wood	Preparation of three sections: transverse, radial and tangential. Mounting in glycerin alcohol or permanent medium and staining. Even slivers and sawdust can be sectioned.	Presence or absence of bordered pits, crossfield pits. Recognition of arrangements of tracheids or fibers and vessel elements. Detailed study of pitting types on vessels and tracheids. The arrangement of the cells in all three sections. Comparison to reference slides and illustrations.
Pollen	Acetolysis (boiling in mixture of acetic anhydride and sulfuric acid) to remove cytoplasm. Mounting in glycerin, glycerin jelly or silicon oil.	Shape. Type and arrangement of apertures: pores and/or furrows. Structure (e.g. collumellae) and sculpturing (e.g. echinate) of the exine. Comparison to reference slides and atlas figures.
Diatoms	Boiling in strong acids to destroy organic matter. Mounting in high refractive index medium.	Shape of girdle and valve views, size, arrangement of pores, fine structure of the central raphid in the center of the diatom.
Blood crust	Reconstitution in reagent such as toluidine blue.	Recognition of leukocytes and other cells which may give clues as to other tissues present.

locate vibration directions) and by convergent polarized light (to determine the optical character, optic sign and orientation).

In addition to the microscope itself, a set of calibrated index of refraction oils are required to perform optical crystallographic measurements. Five basic optical properties may be determined using this equipment. These are the refractive index (for an isotropic solid) or refractive indices (for an anisotropic solid), birefringence, optic sign, interference figure and pleochroism. **Table 2** summarizes the determinative methods used along with some typical applications.

Almost any type of transparent solid particle can be identified on the basis of its optical properties by a trained polarized light microscopist. Typical examples include: mineral grains from sand and soil, synthetic and regenerated fibers, drug crystals, building materials, cosmetics, automotive and architectural paint pigments and extenders, explosives and dust. Tentative identifications performed by polarized light microscopy can be confirmed by microchemical analysis using classical or instrumental methods when appropriate. In certain cases, the observation or measurement of these optical properties may also act as points of comparison. Two examples of this are the recognition and comparison of specific varieties of minerals in a soil comparison based on pleochroism of the grains and the exact measurement of the birefringence of polyester fibers from a cotton/polyester shirt and fibers recovered from a knife which was alleged to have cut it. In the first example, the distinctive pleochroic colors of the mineral grains identify them as different varieties of the same mineral, a characteristic which can make a comparison more certain or less probable, depending on their presence or absence in the known and questioned soils. In the second example, the microscopist could measure the exact birefringence of the polyester fibers from the knife and compare them to the fibers from the shirt, if measurements from the fibers comprising the shirt all had the same value.

The *comparison microscope* is used to compare microscopic items side by side. Although the human eye can be very good at discerning minute differences in color and morphology, the brain has a more difficult time remembering and processing these subtle differences. This problem is overcome by a comparison microscope in which the images from two microscopes are observed side by side in a single field of view. Reflected light instruments are used by firearms examiners to compare rifling marks on bullets as well as ejector marks and firing pin impressions on cartridge cases. Tool marks and cut and polished layered paint chips can also be compared with the same equipment. Transmitted light microscopes are

Table 2 Optical properties and their determination with the polarizing microscope

Optical property	*Measurement*	*Determinative value*
Refractive index	Immersion method. Specimen placed in index of refraction oils until it disappears.	Identification of unknown isotropic crystals. Comparison of glass particles.
Refractive indices	Orientation of principal vibration direction with polarizer of microscope. Matching of two (uniaxial crystals) or three (biaxial crystals with index of refraction oils.	Crystallographic identification of chemical crystals and minerals. Identification and comparison of artificial fibers.
Birefringence	Numerical difference between two principal refractive indices. By subtracting larger refractive index from smaller or by measurement of retardation with a compensator and measurement of thickness.	Identification of crystalline and semicrystalline materials. Comparison of certain artificial fibers.
Optic sign	Convention based on relative magnitudes of refractive indices. By comparison of values of refractive indices or by means of compensators.	Aid in the identification of crystals, minerals and fibers. For artificial fibers the more easily determined sign of elongation is the same as the optic sign.
Interference figure	Viewed in convergent light between crossed polars with objective of high numerical aperture using a Bertrand lens or pinhole.	Aid in determining the optical character of crystals as uniaxial or biaxial. Provides optical orientation of crystals without diagnostic external morphology.
Pleochroism	Rotation of the crystal, particle or fiber between extinction positions with only one polarizer inserted in the optical path.	Diagnostic aid in the identification of heavy minerals and their varieties in soil mineralogy. Comparison aid in the examination of colored artificial fibers.

used to compare hairs, fibers and layered paint chips which have been thin-sectioned. Polarizing and fluorescence equipment may added to a comparison microscope which is to be used for fiber comparisons to enhance its capabilities. Human hair comparisons, particularly the final stages of an examination, are conducted almost exclusively under a comparison microscope.

The *phase contrast* microscope is used primarily in serological and glass examinations. Its principal use in serology is to observe cells in biological fluids or after reconstitution in aqueous mountants. Under these conditions cells exhibit a very small optical path difference with respect to the medium in which they are immersed. Such specimens are referred to as phase objects. The human eye cannot observe phase differences, but it can discern amplitude (dark and light) differences. A phase contrast microscope uses half-silvered rings and disks placed in the optical system to change these phase differences into amplitude differences which can then be observed and photographed. Spermatozoa, epithelial cells, and other cellular matter can be studied in detail without staining using this technique. One of the principal methods of glass comparison is based on a very accurate measurement of the refractive indexes of the known and questioned samples. The measurement is conducted in a hot stage mounted on a phase contrast microscope. The crushed glass fragment is mounted between a slide and coverslip in a specially prepared and characterized silicone oil which is placed in the hot stage. As the temperature is raised, the refractive index of the silicone oil decreases while that of the glass remains essentially constant. Since even small differences between the refractive indexes of the glass and oil are easily seen with phase contrast, the true match point (i.e. temperature at which the silicone oil has the same refractive index as the glass) can be observed with great precision. The refractive index of a glass particle can be measured to 0.00002 using this technique. A commercial instrument in which the phase contrast microscope, hot stage and camera are all connected to a computer makes these measurements automatically and objectively.

Fluorescence microscopy is based on the property of certain substances to emit light of a longer wavelength after they have been irradiated with light of a shorter wavelength. This emitted light is called fluorescence and differs from luminescence in that the emission of light stops after the exciting radiation is switched off. The fluorescence may originate from fluorescent 'tags' attached to proteins or other compounds which cause the substance they react with to fluoresce after the nonreacting remainder of the reagent is washed away, or it may originate from autofluorescence. The first technique is the basis for detecting antigen–antibody reactions which occur on a cellular level and has been applied to a limited extent in forensic serology.

Autofluorescence may originate from either organic or inorganic compounds or elements. When it occurs, autofluorescence is a useful comparison characteristic. It may originate from organic dyes or optical brighteners on fibers; it may be observed in layers of paint in cross-section where it originates from organic pigments or inorganic extenders and may be observed on certain varieties of mineral grains and be absent from others. A modern fluorescence microscope is equipped with a vertical illuminator which directs the light from a mercury burner through a series of lenses and filters designed to focus the light on the specimen and select a narrow or wide range of wavelengths to excite fluorescence in the specimen. Since the intensity of the fluorescence from a specimen does not depend on absorption and the image is formed with the emitted fluorescent light rays, fluorescence images are bright and well resolved. These images can be recorded and make excellent exhibits for use in reports and courtroom presentations.

The *hot stage* microscope permits the microscopist to observe the behavior of specimens as they are exposed to temperatures from ambient up to approximately 350°C. Melting temperatures can be used to help in the identification of unknown substances and as an aid in certain types of comparisons; particularly those involving thermoplastic polymers. For example, infrared microspectroscopy is of only limited use in distinguishing nylon fibers. It can be used to determine if fibers have been spun from nylon 6 or nylon 6,6 polymer. Much finer distinctions can be made by comparing melting points of nylon fibers since these are a function not only of the type of polymer from which the fiber was spun, but also the average molecular weight, crystallinity, presence of additives, etc. Although the contributions from each of these factors cannot be individually assessed from a melting point determination alone, the actual melting points of two fibers result from all of these factors and thus form a useful point of comparison or discrimination. Although other instrumental methods of analysis have largely superseded hot stage microscopy as a tool for the identification of unknown compounds, it is still a useful technique which can add information and make distinctions which are difficult or impossible by other methods. The identification of polymorphs of drugs of abuse, for example, is better studied by thermal methods than by spectroscopic ones. Determination of the melting range can also give information on the purity of a minute sample which could be difficult to assess by other means.

Electron microscope

Electron microscopes make use of electrons rather than photons to form their image. The transmission electron microscope (TEM) was developed first, followed some years later by the scanning electron microscope (SEM). Transmission instruments are generally more difficult to use and require more painstaking sample preparation than scanning microscopes and thus have found very few applications in forensic science. Specimens for TEM must be extremely thin to permit penetration by the electron beam. The image in an SEM is formed from collected secondary or backscattered electrons emitted from (and just beneath) the surface of the sample and not by transmitted electrons as in the TEM. Since the SEM only looks at the surface of a specimen, sample preparation is often much simpler and frequently consists simply of placing the specimen on a piece of conductive carbon tape. It may be necessary to vacuum deposit a layer of carbon or gold over nonconductive specimens to make them conductive, although the new 'environmental SEMs' can image nonconductive samples in a low vacuum. SEMs are now in use in many forensic laboratories around the world. Most of these microscopes are equipped with energy dispersive X-ray spectrometers for elemental analysis. X-ray spectrometers collect the X-rays which are produced along with the secondary and backscattered electrons when a specimen is bombarded in a vacuum with electrons. These X-rays are collected and then sorted in a multichannel analyzer according to their energy which is directly related to atomic number. Both qualitative and quantitative analyses can be performed on microscopic specimens from boron all the way up in the periodic table. The detection limit for each element varies, but typical limits of detection for most elements, excluding some of the light elements, is about 0.1%.

One of the principal uses of analytical SEMs in forensic science laboratories is the detection and analysis of gun shot residue (GSR) particles. Conductive sticky tape, attached to the back of a sample stub, is pressed over a suspect's hands to collect any residue which might be present. The stub is placed in the microscope and searched, either manually or automatically, for particles with a spherical morphology which contain lead, antimony and barium. The combination of the spherical morphology with this elemental composition provides better proof of the presence of GSR than an elemental analysis alone. Other types of microscopic evidence which can be examined in the SEM include items as diverse as pollen grains, diatoms, paint, glass, inorganic explosives and general unknowns. The combined abilities of the SEM to resolve fine structures and provide the elemental composition of these small particles is a tremendous aid in the examination of many small items of trace evidence.

Forensic Microscopy

Although most forensic scientists use microscopes at one time or another, the forensic microscopist uses microscopes to locate, recover, identify and compare trace evidence on a daily basis. It is essential that these scientists be trained in the use of the microscope as an analytical tool. Thus, they must understand the geometrical optics essential for image formation and the physical optics which govern resolution. They must have learned polarized light microscopy and optical crystallography and mineralogy in order to identify unknown crystalline materials and artificial fibers and to compare sands and soils. Microchemical analysis, using both classical and instrumental methods, is essential for the study of materials which can be compared by means of their elemental and/or chemical composition. The forensic microscopist must also learn the essential identification features of human and animal hairs, vegetable fibers, pollen, diatoms, wood and plant anatomy in general. These substances cannot be identified or compared by chemical analysis; only morphological characteristics distinguish one species or variety of these natural materials from another. In this regard, the microscopist must also become proficient at preparing samples properly for the instrument which will be used to make the test or observation and in the interpretation of the results.

Because of the small size of the samples usually available for analysis, the microscopist must learn to optimize the two analytical parameters that are under his/her control. The first is the microscope. The proper instrument must be selected for the measurement or observation to be made and it must then be adjusted so that it optimally performs this function. The sample must then be prepared to minimize artifacts and maximize the amount of useful information which can be obtained with due regard to preservation of as much of the sample as possible, but also with the recognition that the importance of the evidence lies not in its preservation but in the factual information which it may provide to the investigation. In comparisons, it is essential that both the known and questioned samples be prepared in an identical manner.

See also: **Fibres**: Identification and Comparison. **Hair**: Comparison: Microscopic. **Wood Analysis**.

Presumptive Chemical Tests

B Levine, Office of the Chief Medical Examiner, Baltimore, MD, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0412

Introduction

Different types of evidence may be presented to a forensic chemistry laboratory for analysis. This includes solid or liquid material such as powders, tablets, capsules, plant material or solutions and paraphernalia such as syringes, cookers, spoons, cigarettes and pipes. Although some of the components of the evidence may be suggested by external appearance, the forensic chemist is still required to perform analytical testing on the submitted evidence. The material may contain 'active' drug or drugs as well as inactive substances used to dilute the potency of the preparation. As workload for the forensic chemist increases, it would be impractical to subject all submitted evidence to sophisticated and time-consuming extraction and instrumental techniques. Therefore, there is a need for a battery of relatively simple screening tests that the chemist can perform to assist in directing toward an ultimate identification.

The forensic toxicologist frequently faces the same challenges that the forensic chemist encounters. For instance, in postmortem cases, death may occur from exposure to one or more toxic substances. Most of the cases involve alcohol and therapeutic or abused drugs, but other cases may involve gases, volatile substances, pesticides, metals or environmental chemicals. The postmortem forensic toxicologist must be equipped to assist the medical examiner or coroner in the investigation of death due to any of the above substances. In a 'general unknown' case, most toxicology laboratories perform a series of tests designed to screen for a large cross-section of substances that would reasonably be encountered. Many of these tests are designed to detect or rule out a class or classes of drugs.

One of the oldest analytical procedures used by forensic chemists or forensic toxicologists is broadly classified as 'color tests'. Color tests or spot tests are chemical tests which involve the reaction of a sample with a reagent or a series of reagents to produce a color or change in color. The biggest advantages to color tests are their simplicity and ease of use. No sophisticated equipment is required and the time needed to train analysts is short. A negative result for a color test is helpful in excluding a drug or a class of drugs, depending on the test performed. It is interesting that many of these color tests were developed empirically. The chemistry behind these color tests is often complex or unknown.

A chemical test with greater specificity than the color test is the microcrystalline test. To perform a microcrystalline test, a drop of reagent is added to a small quantity of sample on a microscope slide. After the reaction is completed, crystals with unique size and shape develop in relation to the substance present. Although these tests may be performed directly on the powder, the presence of dilution agents may change the structure of the crystals.

It must be emphasized that all chemical tests must be confirmed by an alternate analytical technique based on a different chemical principle. In the forensic chemistry laboratory, the confirmatory test usually includes infrared spectrophotometry or mass spectrometry. In the forensic toxicology laboratory, gas chromatography/mass spectrometry is currently the method of choice for confirming the presence of most drugs.

This article is divided into two sections; the first section deals with chemical tests that are often used or are otherwise available to the forensic chemist to test evidence such as material or paraphernalia. The second section discusses chemical tests employed by the forensic toxicologist to screen biological specimens for toxic substances. It is beyond the scope of this article to include all possible chemical tests available to the forensic chemist or toxicologist. Instead, many of the currently used or historically important tests are covered.

Chemical Tests on Evidence

Color tests

There are a large number of color tests that have been developed over the years to screen for the presence of certain drugs or drug classes. The following are some of the more common color tests.

Dille–Koppani test Two reagents are required for this reaction: 1% cobalt nitrate in methanol and 5% isopropylamine in methanol. After sequential addition of the two reagents, a violet color is obtained with imides in which the cabonyl and amine are adjacent in a ring. This would include barbiturates, glutethimide and saccharin. No color reaction occurs when there are substituent groups on the nitrogen atom.

Duquenois–Levine test Two reagents and chloroform comprise this reaction. The first reagent added is a mixture of 2% vanillin and 1% acetaldehyde in

ethanol; the second reagent is concentrated hydrochloric acid. A blue–purple color indicates cannabis, coffee or tea. Cannabis may be differentiated from the other substances by extracting with chloroform. If the purple color enters the chloroform layer, then the test is presumptively positive for cannabis.

Erlich test The addition of 1% *p*-dimethylaminobenzaldehyde in 10% hydrochloric acid in ethanol to a sample containing ergot alkaloids yields a violet color. This serves as a rapid screening test for LSD.

Liebermann test The reagent is prepared by adding 5 g of sodium nitrite to 50 ml sulfuric acid. A number of colors, from orange to orange–brown to yellow are produced. Different structures are suggested by the color. The test needs to be repeated with sulfuric acid alone since many compounds produce color changes with the acid.

Mandelin test This is another reaction that produces a wide array of colors and must be interpreted in light of the Liebermann test. The reagent is 0.5% ammonium vanadate in sulfuric acid.

Table 1 lists the colors produced by common drugs with both the Liebermann and the Mandelin tests.

Marquis test This reagent is prepared by mixing one part formaldehyde to nine parts sulfuric acid. By adding one drop of this reagent to the sample, a multitude of colors may arise. Structures that give violet colors include a ring sulfur, a ring oxygen or aromatic compounds consisting entirely of carbon, hydrogen and oxygen. Heroin, morphine and most opioids produce a violet color. Amphetamine and methamphetamine produce an orange–brown color with this reagent. **Table 2** lists a number of drugs and the colors generated with this reagent.

Table 1 Color reactions of some common drugs with Liebermann's and Mandelin's reagents

Drug	*Color*	
	Liebermann	*Mandelin*
Amphetamines	Red–orange	
Codeine	Black	
Diphenhydramine	Brown–orange	Yellow
Glutethimide	Red–orange	
LSD		Gray
Meperidine	Red–orange	
Mescaline	Black	
Methadone	Brown–orange	Green to blue
MDA	Black	Green to blue
Morphine	Black	Blue–gray
Phenmetrazine	Red–orange	
Propoxyphene	Brown	Gray–brown
Psylocibin		Green

LSD, lysergic acid dimethylamide; MDA, methylenedioxyamphetamine.

Scott test This is a series of color reactions used to screen for the presence of cocaine. Two percent cobalt thiocyanate in water and glycerine (1:1) will cause a powder containing cocaine to turn blue. The addition of concentrated hydrochloric acid changes this color to a clear pink. The blue color reappears on extraction with chloroform.

Microcrystalline tests

It is recommended that a polarizing microscope be used for the microcrystalline tests. The best magnification for observing crystals is approximately 100 ×. The description and classification of microcrystals can be a complex process. When a reagent is added to different substances, differences in shape, color or dichroism may result. The following is a general description of some of the shapes that may be encountered when performing microcrystalline tests. These descriptions are based on the classification of Fulton (1969) in *Modern Microcrystal Tests for Drugs*.

- Bars: Solid appearing crystals with three unequal dimensions.
- Blades: Thin flat crystals with much greater length than width.
- Grains: Granular precipitate without distinction at 100 × magnification, or crystals that have approximately the same length, width and thickness.
- Needles: Crystal with little width or thickness in comparison to length.
- Plates: Thin flat crystals with width comparable to length.
- Rods: Solid appearing crystals with greater length than width or thickness; width and thickness significant and equal in magnitude.

Table 2 Drugs that react with the Marquis reagent

Drug	*Color*
Amphetamine	Orange–brown
Aspirin	Pink
Codeine	Purple
Diphenhydramine	Yellow
Fentanyl	Orange
Heroin	Purple
Meperidine	Orange
Methamphetamine	Orange–brown
Morphine	Purple
Opium	Purple
Propoxyphene	Black
Psilocybin	Yellow

- Rosettes: Aggregate of crystals that grow out from a central point in all directions.
- Tablets: Flat crystals with similar length and width greater than thickness

Beside a general shape, crystals will often have a characteristic angle of form that may be measured. In addition, crystals will often display dichroism or two different colors with different orientation of polarized light. Although additional colors may be seen, they are usually gradations of the two extreme colors. Undoubtedly, the best description of crystal forms for a particular test is via a photograph. Furthermore, the unknown crystals should be compared to crystals obtained with actual standard material.

Another factor in the type of crystals formed is the reagent or reagents used to generate the crystals. **Table 3** lists some of the most common reagents used. The solvent in which these reagents are dissolved is also a significant factor in the types of crystals produced. The general medium for these tests is water. Several aqueous acids have been used most commonly as solvents, including phosphoric acid, sulfuric acid, hydrochloric acid and acetic acid. These acids alter the solubility characteristics of the crystals. For example, phosphoric acid produces the greatest insolubility whereas acetic acid yields the greatest solubility.

Table 3 Common microcrystalline reagents

Chemical	*Symbol*
Bromauric acid	$HAuBr_4$
Chlorauric acid	$HAuCl_4$
Chloroplatinic acid	H_2PtCl_6
Iodobismuth acid	H_3BiI_6
Mercuric chloride	$HgCl_2$
Mercuric iodide	HgI_2
Potassium iodide	KI
Potassium permanganate	$KMnO_4$
Silver nitrate	$AgNO_3$

Although the use of microcrystalline tests in forensic chemistry has diminished over the years with the availability of more advanced instrumentation, there are still a number of tests which are still in use for commonly seized substances. **Table 4** lists a common test for some of these substances.

Chemical Tests on Biological Specimens

Some biological specimens are amenable to presumptive chemical tests without any specimen pretreatment. For instance, color tests can often be performed directly on urine or stomach contents. Other specimens, such as blood, bile, liver and kidney require some type of pretreatment prior to the chemical test. Although liquid–liquid extraction or solid-phase extraction can precede the chemical test, there are simpler separation techniques that are more 'appropriate' given the presumptive screening nature of these tests. Two such separation techniques are Conway microdiffusion and protein precipitation.

A Conway microdiffusion cell is a porcelain dish containing either two or three concentric reservoirs with an accompaning glass cover. A trapping reagent is added to the center well. As the name implies, the trapping reagent captures the analyte of interest for the subsequent chemical reaction. In some situations, the trapping reagent is the chemical reagent. In a two-reservoir well, the blood or tissue specimen is added to the outer well, followed by the releasing agent. This releasing agent causes the analyte to leave the biological sample. A glass cover is rapidly placed over the cell, which is then swirled for a brief time. The reaction then proceeds at room temperature for several hours. The glass cover is usually greased to prevent release of analyte from the cell. Alternatively, a three

Table 4 Examples of microcrystalline tests for commonly abused drugs

Drug	*Reagent(s)*	*Crystals*
d-Amphetamine	5% $HAuBr_4$ in (1 + 2) H_3PO_4 + 1(2 + 3) H_2SO_4	Light colored trapezoidal blades
Cocaine	(1) 20% HOAc (2) 5% $HAuCL_4$ in H_2O	Long rods with short arms at right angles
Heroin	5% $HgCl_2$ in H_2O	Rosettes of needles
d-Methamphetamine	H_3BiI_6 in (1 + 7) H_2SO_4, hanging drop	Orange needles
Phencyclidine	10% KI in water	Branching needles

H_3PO_4, phosphoric acid; H_2SO_4, sulfuric acid; HOAc, acetic acid.

well cell is used. The sample and releasing reagent are added to the middle well while a sealing agent is added to the outer well. After the reaction is complete, the center well contains the trapped analyte.

Protein precipitation is another simple pretreatment technique. The protein content of human body fluids and tissues is considerable, from about 6% by weight in plasma to greater than 50% by weight in liver and other organs. Numerous reagents have been developed to precipitate protein; two that are commonly used in drug analysis are trichloroacetic acid (10–15% in water) and tungstic acid (10% sodium tungstate in water and used in conjunction with 3N sulfuric acid). Once proteins have been precipitated, separation of aqueous and solid protein must occur by filtration or centrifugation.

Common color tests: classes of substances

A number of color reactions have been devised to identify a number of drugs within a particular drug class. Many of the color reactions on drug classes are used in conjunction with thin layer chromatography. Once the individual substances have been separated on the thin layer plate, these color reagents are applied to the plate for visualization of color. Interpretation of color tests for drug classes must be done with caution. Drugs within a class may have therapeutic ranges that differ by orders of magnitude. This means that a negative result may not exclude the presence of a drug whereas a positive test does not necessarily mean that a toxic amount of drug is present.

Barbiturates Mercuric nitrate and diphenylcarbazone will react with a chloroform extract of a biological specimen to produce a purple color. The specific barbiturate is not identified by this method. Phenytoin will also produce a purple color.

Benzodiazepines The Bratton–Marshall test is a classic screening method for identifying benzodiazepines. Benzodiazepines are converted into benzophenones by acid hydrolysis and heating. After an extraction, the following color reagents are added: 10% sulfuric acid, 0.1% sodium nitrite, 0.5% sulphamic acid and 0.1% *N*-1-naphthylethylene diamine. A purple color will ensue if a benzophenone is present.

Carbamates After an alkaline extraction and concentration, the residue is treated with 10% furfural in ethanol and exposed to fumes of concentrated hydrochloric acid. A blue–black color indicates the presence of carbamates such as carisoprodal, meprobamate or methocarbamol.

Chlorinated hydrocarbons The Fujiwara test is a classic assay to identify trichlorinated compounds such as trichloroethanol. To urine or a tungstic acid precipitate of blood or tissue is added concentrated sodium hydroxide and pyridine. The mixture is placed in a boiling-water bath for several minutes; the presence of a pink color in the pyridine layer is a positive test. It is especially critical to run a reagent blank when performing the Fujiwara test as some of the reagents used may have contaminants which yield an apparent positive result.

Heavy metals This screening test was developed by the German chemist Reinsch and is still used as a qualitative screening test for an overdose of arsenic, antimony, bismuth, or mercury. The method is based on the fact that copper will displace from solution elements below it in the electromotive series. This method involves boiling a small copper coil in an acidified solution of urine, liver or stomach contents. If any of the four mentioned metals are present in the specimen, they will replace the copper on the coil as a dark film. A black- or purple-colored deposit may be due to arsenic, antimony or bismuth; a silver or gray deposit may be due to mercury. Sulfur compounds can interfere with this test by producing a sulfide salt. This method lacks sensitivity but can usually determine an overdose.

Phenothiazines In the presence of an appropriate oxidizing agent, phenothiazines will produce a color. One such reagent is known as the 'FPN' reagent and is a combination of ferric chloride, perchloric acid and nitric acid. A variety of colors may be produced, depending on the phenothiazine present. This test may be useful in the identification of an intoxication of some phenothiazines.

Common color tests: specific substances

Acetaminophen The test for acetaminophen is performed on urine or a protein-free filtrate of blood and requires heating at 100°C after the addition of hydrochloric acid. A blue color after the addition of 1% *o*-cresol in water and ammonium hydroxide constitutes a positive test for acetaminophen. Therapeutic or toxic use of aceteminophen can be identified using this color reaction.

Carbon monoxide (CO) One of the early methods of CO analysis involved microdiffusion using a Conway cell. The specimen is placed in the outer well and sulfuric acid is added to release the CO from hemoglobin. A solution of palladium chloride is added to the center well. The cell is sealed and incubated at

room temperature for one to two hours. As the reaction proceeds, the palladium chloride is reduced to metallic palladium, forming a black or a silver mirror in the center well and the CO is changed to carbon dioxide. This method is capable of distinguishing between normal carboxyhemoglobin saturation levels ($<10\%$) and elevated carboxyhemoglobin saturation levels.

Cyanide In a two-reservoir well of a Conway microdiffusion cell, the blood or tissue specimen is added to the outer well, followed by the releasing agent. This releasing agent may be a dilute mineral acid such as sulfuric acid or may be an organic acid such as tartaric acid. The releasing agent causes the formation of the gaseous hydrocyanic acid (HCN). To the center well is added dilute base. This serves as a trapping reagent for the released HCN. The glass cover is rapidly placed over the cell, which is then swirled for a brief time. The reaction then proceeds at room temperature for several hours Alternatively, a three-well cell is used. The sample and acid are added to the middle well and a more dilute acid is added to the outer well. This acts as a sealing agent. After the reaction is complete, the center well contains the trapped cyanide and is available for detection.

A number of colorimetric reactions have been developed to detect cyanide from the trapping agent. One classical color reaction uses chloramine-T to convert cyanide in to cyanogen chloride. A color reagent containing pyridine and barbituric acid is then added to produce a red color. Another common color reagent uses *p*-nitrobenzaldehyde and *o*-dinitrobenzene which, when added to the trapping agent produces a violet color.

Ethanol Most of the early methods for ethanol analysis in biological fluids and tissues used wet chemical methods. Due to its volatility, ethanol can be easily separated from a biological matrix by distillation or microdiffusion. Microdiffusion is more amenable to batch analysis than is distillation. Potassium dichromate and sulfuric acid is a common color reagent which is placed in the center well of a Conway microdiffusion cell. In the presence of ethanol or any other volatile reducing substance, dichromate is converted to the chromic ion, with a change of color from yellow to green. This color test can be observed visually, making this a simple screening method. The main drawback to this method is that it is nonspecific for ethanol; other aldehydes and ketones will oxidize and interfere with the ethanol quantitation.

Ethchlorvynol Although ethchlorvynol is rarely used today as a sedative–hypnotic, a specific and sensitive color test had been developed during a period of greater use. Diphenylamine is added to urine or a protein-free filtrate of blood. Concentrated sulfuric acid is gently poured down the side of the tube and a red color at the interphase represents a positive test.

Imipramine The Forrest reagent can be used to identify toxic concentrations of imipramine, desipramine, clomipramine or trimipramine. The reagent consists of equal portions of 0.2% potassium dichromate, 30%, v/v, sulfuric acid, 20% perchloric acid and 50% nitric acid. A green color is a positive test for any of the mentioned drugs. Phenothiazines will also produce a color with this reagent.

Methanol A color test specific for methanol involves the oxidation of methanol to formaldehyde by 3% potassium permanganate. Sodium bisulfite and chromotropic acid are then added, followed by a layering of concentrated sulfuric acid. A purple color at the acid/filtrate interface is a positive test. The test needs to be repeated without the addition of permanganate to rule out the presence of formaldehyde in the specimen.

Nitrite Although nitrites have been used for many years as antihypertensive agents, the need to test for nitrites in urine specimens has taken on greater importance over the past several years. Nitrite adulteration causes a significant reduction in the recovery of the major urinary metabolite of marijuana when tested by gas chromatography/mass spectrometry. Urine specimens can be screened for the presence of nitrites by adding 0.6% sulfanilic acid in 20% hydrochloric acid and 0.48% naphthylamine in 20% hydrochloric acid. A red color is a positive screening test.

Paraquat Paraquat is reduced by an alkaline solution of 1% sodium dithionite to produce a purple color. Diquat will yield a green color under similar conditions.

Salicylate Salicylate, the metabolite of aspirin, reacts with an acidic solution of ferric chloride to produce a purple color. This color reaction requires the presence of both the free phenolic group and the free carboxylic acid group that appears on the salicylate molecule. Therefore, aspirin itself will not produce a positive result prior to hydrolysis to salicylate. This color test has sufficient sensitivity to detect therapeutic use of salicylate.

See also: **Drugs of Abuse**: Classification, including Commercial Drugs. **Toxicology**: Overview; Methods of Analysis – Ante Mortem; Methods of Analysis – Post Mortem; Interpretation of Results. **Pharmacology**.

Further Reading

Baselt RC (1981) *Analytical Procedures for Therapeutic Drug Monitoring and Emergency Toxicology*. Davis, CA: Biomedical Publications.
Feigl F (1966) *Spot Tests in Organic Analysis*, 7th edn. Amsterdam: Elsevier.
Fulton CC (1969) *Modern Microcrystal Test for Drugs*. New York: Wiley.
Moffat AC, Jackson JV, Moss MS and Widdop B (eds) (1986) *Clarke's Isolation and Identification of Drugs*. London: The Pharmaceutical Press.
Saferstein R (ed.) (1998) *Criminalistics: An Introduction to Forensic Science*, 6th edn. Upper Saddle River, NJ: Prentice Hall.
Sunshine I (ed.) (1975) *Methodology for Analytical Toxicology*. Cleveland, OH: CRC Press.

Separation Techniques

C Bommarito, Michigan State Police Forensic Science Division, East Lansing, MI, USA

Copyright © 2000 Academic Press

doi:10.1006/rwvi.2000.0409

Many types of evidence (drugs, paint, explosive residue, fire debris, soil, biological fluids, etc.) encountered in the forensic science laboratory consist of complex mixtures of substances. The complexity of these materials is a double-edged sword to the forensic scientist. The more complex and variable a mixture, the greater its probative value when comparing known and questioned samples. Complex mixtures also create analytical problems, as most compounds need to be relatively pure in order to be identified by analytical techniques, such as spectroscopy. It is the ability of the forensic chemist to separate that allows him/her to identify or compare a material. Separation techniques discussed in this article can be classified into four groups: physical, chemical, chromatographic and electrochemical.

Physical Separations

Physical separations are commonly used in forensic chemistry. Examination of a mixture under a stereoscope may yield variation of its component particles. These particles may vary in their shape, size, color, opacity, texture or other physical properties, which can be observed microscopically. The particles may then be physically separated from the mixture. Particle selection or 'picking' is the major separation technique used in most areas of trace evidence. A skilled examiner can recognize and pick out critical particles in explosive residue. Paint examiners can select uncontaminated particles out of a mixed smear of paint for examination and comparison. This type of separation can also be utilized in mixtures containing cocaine hydrochloride. This salt form of cocaine has a translucent shale-like appearance, which is easily distinguished from the coarser grains of most diluents and excipients used as cutting agents in cocaine mixtures.

Forensic geologists have long used particle size to separate soil components for analysis. A novel example of this is the use of a mesh screen for separation of sand grains onto an SEM stub for automated elemental analysis and mineral identification (**Fig. 1**). Manual separation of particles should always be explored prior to attempting a chemical separation.

Volatile Materials

Volatile materials are often present in drug and fire debris cases submitted to forensic laboratories. These materials (primarily solvents and flammables) may be easily separated from relatively nonvolatile substances by techniques such as distillation, sublimation, headspace analysis, absorption–elution and solid-phase microextraction.

Distillation and sublimation are techniques occasionally used in the analysis of drugs. Amines such as amphetamine and methamphetamine are relatively

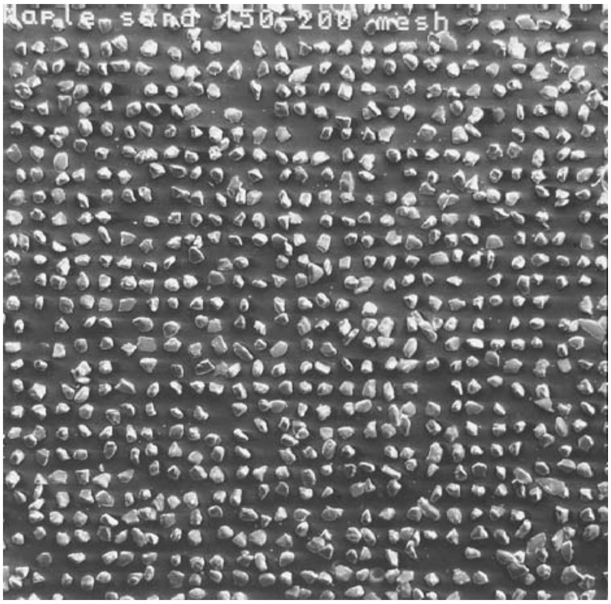


Figure 1 Photograph of sand particles separated using a mesh screen and mounted on a SEM stub. (Courtesy M.J. McVicar and W.J. Graves, Ontario Centre of Forensic Sciences.)

volatile and may be flash distilled to separate the drug from a complex mixture. Dimethylsulfone is currently one of the most common cutting agents used with methamphetamine in the United States. It is far more volatile than methamphetamine, sublimating at 90–100°F (35–38°C). It can be easily removed from methamphetamine by placing the mixture on a watch glass and sublimating the dimethylsulfone over a steam bath. The purified methamphetamine can then be analyzed via infrared spectrometry.

Examination of fire debris for the presence of ignitable liquids (gasoline, fuel oil, etc.) requires that the residual amounts of these liquids be separated from the debris. Samples of debris are sealed into airtight nylon bags or paint cans and heated to volatilize the ignitable liquid into the headspace of the container. The volatile in the headspace can then be analyzed directly by removing a portion with a syringe. Often the concentration of the volatile in the headspace is insufficient for direct analysis. In these cases, a process known as absorption–elution may be used to concentrate the volatile. In this process, the volatiles are absorbed by a material (usually activated charcoal) and then extracted from the absorbent in a concentrated form by a solvent. The process of absorption can be active or passive. In active or dynamic absorption, the material containing the volatile is heated and the vapor forced through a vessel containing the absorbent. A typical setup may consist of the charcoal housed inside a disposable pipette or hypodermic needle connected to a vacuum pump. The open end of the vessel is placed in the headspace of the sample and the air containing the ignitable liquids is drawn through the vessel. The charcoal would then be removed for solvent extraction of the captured volatiles (elution). Passive absorption entails placing the absorbent material directly inside the airtight container while heating to volatilize the ignitable liquids. Vaporized molecules of the volatiles contact the absorbent in a random fashion. Because the volatiles are not forced to contact the absorbent, the time of exposure of the absorbent to the vapor must be dramatically increased in order to ensure proper absorption.

Solid-phase microextraction (SPME) is a relatively new technique that has been used increasingly in forensic work. In this technique, a special syringe containing a spring-loaded fiber coated with a bonded phase is used. The fiber in the syringe can be contracted and retracted. The sample containing the volatiles is sealed in an airtight container. The syringe is then inserted into the headspace of the container and heated for a period of time. The volatiles are adsorbed onto the fiber, which is then retracted into the syringe. The syringe is then used for injection in a gas chromatograph. The fiber is contracted from the syringe into the injector port and the heat of the injector port elutes the volatiles from the fiber (**Fig. 2**). SPME may also be used to concentrate analytes from a liquid solution onto the SPME fiber. This technique is increasingly being used in the areas of solvent, drug, fire debris and high explosives analysis.

Chemical Separations

Solvent extraction

Although chemical separations have increasingly become chromatographic in nature, solvent extraction is still frequently used in the forensic laboratory, especially in drug analysis, where chemical separations are generally followed by confirmatory qualitative techniques such as infrared spectrometry. Solvent extractions can be either single-phase or multiphase in nature. The theory and process of both types of solvent extractions are based on many criteria, the most important of which is solubility.

In a single-phase extraction, components of a mixture are separated based on the solubility of the components in a solvent. In this type of extraction, solvent selection is critical. The target component of the mixture must either be soluble in the solvent, while all remaining components are insoluble or vice-versa. The solvent is added to the mixture and then filtered to separate undissolved solute and solvent. If the target compound is soluble in the solvent it is then recovered by evaporation of the solvent. If the target compound is insoluble in the solvent and the other components are soluble, the filter paper is then dried

Figure 2 (see color plate 8) Photograph of a SPME fiber inserted into headspace of a sampling vial. (Courtesy R Mindrup, Supelco Inc.)

and the material recovered. This process is also known as 'solvent washing'. In complex mixtures, it is usually not possible to find a suitable single solvent to separate all components of the mixture. If this is the case, several solvents are selected and applied in a serial manner to effect the separation. The selection of solvent generally proceeds from the solvents in which few compounds are soluble (hexane, ether) to those in which a moderate number are soluble (acetone, methylene chloride, chloroform) to those in which most substances are soluble (alcohols, water). Use of acidified or basified solvents (ANOR) may increase the specificity of the extraction.

In multiphase or liquid–liquid extractions, components of a mixture are separated based primarily on their solubility in two immiscible solvents. This technique is also sometimes referred to as partition chromatography. The mixture to be separated is added to a solvent (generally aqueous) and a second solvent (usually organic) is then added. The components of the mixture then partition into the solvent or solvents in which they are most soluble. This type of extraction is extremely flexible, as solvent components can be selected, mixed, their pH changed, heated or cooled, used in vastly different proportions, etc. to effect the desired separation. The most commonly modified variable in multiphase extractions is the pH of the aqueous solvent phase, which is varied by addition of strong acids or bases. The solubility of components dissolved in this phase is directly affected by these pH changes. By adjusting the pH of the aqueous phase to make a particular analyte relatively insoluble in the aqueous phase, the stage is set for the analyte to be extracted into the second solvent, which is then added. Drugs in particular are amenable to separations based on changes of pH in multiphase extractions, lending to a terminology based on their behavior on extraction. 'Acidic drugs' are relatively insoluble in acidic solutions and may generally be extracted from acidic aqueous solutions into an organic solvent. 'Basic drugs' are relatively insoluble in basic solutions and may generally be extracted from basic aqueous solutions into an organic solvent. 'Neutral drugs' are generally unaffected by changes in pH and they will usually stay in whichever phase they are more soluble. An example of such an extraction is the separation of a mixture of codeine, butalbital and acetaminophen. Butalbital is an acidic drug; codeine and acetaminophen are basic drugs. The drug mixture is added to some acidic water, mixed and methylene chloride added. The butalbital will partition into the organic phase, which can then be drawn off and evaporated to recover the drug. The codeine and acetaminophen remain in the aqueous phase, which is then made basic with the addition of a strong base, such as NaOH. Additional methylene chloride is then added. Although acetaminophen is a basic drug, it is mostly insoluble in methylene chloride; therefore, most will remain in the aqueous phase. The organic phase containing the codeine and a small amount of acetaminophen is drawn off and additional basic water added. The remaining acetaminophen is partitioned into the aqueous phase, in which it is more soluble, and the organic phase can then be drawn off and evaporated to recover the codeine.

Purification via chemical reaction

Another chemical technique used in forensic work is purification by salt formation or derivatization. Both techniques involve the addition of a chemical to react with the target component to form a complex, which has properties allowing it to be separated from the rest of the mixture. An example of this is the separation of inorganic acids from a solution. These types of liquids are often encountered in poisoning cases (e.g. a suspect put an acidic solution in a victim's drink) or in clandestine laboratories where acids are utilized in the production of illicit drugs. Knowing that an acid plus a base yields a salt plus water, concentrated ammonium hydroxide is added to the acidic solution, which forms an ammonium salt with the acid. As these salts are insoluble in acetone the addition of this solvent precipitates the ammonium salt from solution. Infrared analysis of the ammonium salt allows the analyst to determine the composition of the acid originally in solution. The production of ammonium chloride indicates that hydrochloric acid was present, ammonium sulfate indicates that sulfuric acid was present, etc.

There are hundreds of chemicals used to derivatize compounds, with the added benefit that the complex formed by derivatization is generally more stable than the underivatized compound. An example of this additional stability is the silylation of psilocybin with BSTFA (*N, O*-bis-(trimethylcilyl)-trifluoroacetamide). Psilocybin (the active drug present in psilocybin mushrooms) undergoes thermal decomposition on heating due to the loss of the dihydrogen phosphate ester group; therefore, due to the heat of the injector, gas chromatography–mass spectrometric (GC-MS) analysis produces chromatograms and spectra that are identical with psilocin. Silylation of the psilocybin prior to GC-MS analysis 'locks in' the phosphate group, eliminating its thermal instability on injection.

Separation of enantiomers

Enantiomers (isomers that are mirror images of each other) are traditionally difficult to separate. Enantiomers have identical melting points, boiling points,

density, dissociation strengths, reaction rates, solubilities, etc. The only variable in which they can be distinguished is the direction of their refraction of plane polarized light (optical activity). Products formed by the reaction of enantiomers with optically inactive reagents are identical, as are the rates of reaction. The reaction rate of enantiomers with reagents that are themselves optically active are not identical and may be so different that one isomer does not react at all. Derivatization of enantiomers with optically active reagents may lead to their separation and their isomeric determination. This is usually accomplished by a chromatographic method following derivatization. Reaction with an optically inactive derivatizing reagent may also yield information about the optical activity of a compound, albeit indirectly. An example of this is in the determination of the optical isomer of methamphetamine. Under US Federal sentencing guidelines, D-methamphetamine (+ rotation) and D,L-methamphetamine (racemic) carry tenfold the penalty as L-methamphetamine (− rotation), necessitating the determination of the optical isomer. The methamphetamine can be derivatized with phenylisothiocyanate (PIT). The resultant derivative precipitates with the racemic mixture but not with the single enantiomers; therefore, if the reaction results in precipitation of the derivative, a racemic derivative of D,L-methamphetamine-PIT forms and its chemical structure is confirmed via infrared spectrometry. If no precipitation occurs, standard L-methamphetamine can be added to the PIT–sample complex and if precipitation occurs after the addition, it can be concluded that D-methamphetamine was initially present. If no precipitation occurs after addition of D-methamphetamine, the entire process must be repeated with the addition of standard D-methamphetamine instead of L-methamphetamine. If precipitation occurs upon the addition of D-methamphetamine to the PIT–sample complex, it can be concluded that L-methamphetamine was initially present.

Microchemical tests

Although superseded in many disciplines by instrumental methods, microchemical tests are an excellent example of the use of chemical reactions for the separation and identification of some compounds. Typically microchemical tests are performed by mixing a small drop containing the substance to be tested with a small drop of reagent. After a few seconds to minutes, characteristic crystals of the resulting compound can be observed and identified via polarized light microscopy. An excellent reference for microchemical tests is Chamot and Mason's *Handbook of Chemical Microscopy, Vol. II*, where scores of microcrystal tests are described and can be used to separate and identify components of mixtures. The quickness, specificity and sensitivity of microchemical tests make them ideal for forensic scientists testing low explosive residue. Using a scheme of microchemical tests to determine the anions and cations present enables the analyst to identify the inorganic oxidizers present in the residue. Some of the more useful tests are summarized in **Table 1**.

Chromatographic techniques

Chromatography involves the separation of compounds based on their distribution between a stationary phase and a mobile phase. There are hundreds of chromatographic techniques utilizing various mobile phases, stationary phases and processes of interaction between the two. In all chromatographic techniques, substances are separated based on their affinity for the stationary phase during the interaction with the mobile phase. Adsorption to the stationary phase is usually caused by the attraction between a polar

Table 1 Common microchemical tests for inorganic ions

Ion	*Reagents*
Aluminum	Ammonium fluoride, ammonium molybdate, cesium sulfate
Ammonium	Chloroplatinic acid, iodic acid, uranyl acetate
Arsenic	Ammonium molybdate, cesium chloride, magnesium acetate
Barium	Ammonium bichromate, potassium ferrocyanide, squaric acid
Beryllium	Chloroplatinic acid, potassium oxalate
Calcium	Ammonium carbonate, potassium ferrocyanide, sulfuric acid
Carbonate	Silver nitrate, calcium acetate
Chloride	Silver nitrate, thallous nitrate
Chromium	Lead acetate, silver nitrate, cesium sulphate
Copper	Potassium iodide, potassium mercuric thiocyanate, uranyl acetate
Cyanide	Silver nitrate, ferrous chloride
Fluoride	Sodium fluosilicate
Iron	Potassium ferrocyanide, potassium thiocyanate
Lead	Ammonium bichromate, potassium iodide
Magnesium	Uranyl acetate
Mercury	Potassium bichromate, potassium iodide
Nitrate	Silver nitrate, nitron
Potassium	Chloroplatinic acid, uranyl acetate, perchloric acid
Sodium	Uranyl acetate, zinc uranyl acetate
Strontium	Ammonium carbonate, squaric acid
Tin	Cesium chloride, oxalic acid
Uranium	Thallous sulfate
Zinc	Oxalic acid, potassium mercuric thiocyanate, sodium nitroprusside

group on the stationary phase and a group of opposite polarity on the sample component. The separated compounds are then processed to visualize the separation, either chemically or more commonly via the use of a detector, which gives an electrical signal that can be graphically displayed (a 'chromatogram'). Often chromatographic methods are used in combination with spectrometric detection to allow the identification of the compounds as they are separated.

Column or liquid–solid (LSC) chromatography LSC utilizes common laboratory materials to effect a separation. The LSC column is typically composed of a laboratory pipette filled with the stationary phase; usually pH adjusted celite or alumina. The mixture to be separated is dissolved in a solvent and the solvent added to the column. Additional solvent (mobile phase) is added until the separated compounds are eluted. The rate at which a compound travels through an LSC column is determined by the amount of time that it spends adsorbed on the surface of the stationary phase. The amount of time a compound takes to elute from the column is known as the retention time. Eluted compounds in column chromatography may be recovered and qualitatively examined. Lysergic acid diethylamide (LSD) may be easily separated by this method using an alumina column. LSD blotter paper is first base extracted into methylene chloride. This crude extract generally contains LSD and dyes and other materials present in the blotter paper. The extract is then added to an alumina column. The LSD at this point does not flow down the column but is retained at the top of the alumina. A few drops of methanol are added which neutralizes some of the active sites in the alumina column, allowing the LSD to flow down the column. Additional mobile phase is added to maintain the flow down the column and the column is observed under UV light to visualize the characteristic fluorescent LSD band as it travels down the column. When the fluorescent band reaches the point of elution, the solvent is collected in a mortar. Potassium bromide is then added and the mixture dried and pressed into a pellet for analysis by infrared spectrometry. Because column chromatography depends on relative adsorption to the stationary phase, this method is also referred to as adsorption chromatography.

Planar chromatography This technique is a form of liquid–solid chromatography in which the stationary phase is held on a plane rather than in a column. The plane may consist of a glass plate coated with the stationary phase (thin-layer chromatography) or a piece of filter paper or cellulose (paper chromatography). Both techniques are commonly used in the forensic laboratory as a screening method and in the analysis of drugs and inks. Thin-layer plates are commercially available with a wide variety of stationary phase coatings to suit most applications. As with other chromatographic techniques, the compounds are separated based on their affinity to the stationary phase. The more tightly bound a substance is to the stationary phase, the less distance it travels up the plane. This rate of travel of the compound in relation to the rate of travel of the mobile phase (R_F) may be used to aid in the identification of the compound. When compared to the travel of a standard under the same conditions, the distance traveled can be measured and if equivalent it is indicative the sample and the standard could be structurally similar. In planar chromatography, the mixture of components to be separated is first dissolved in a small amount of suitable solvent. It is best to select the most volatile solvent in which the mixture is soluble in order to reduce spot size. A portion of the dissolved material is spotted onto the stationary phase with a micropipette. The plate or paper is then placed into a jar containing the mobile phase, which is at a level slightly below that of the spotted samples. The mobile phase travels up the planar surface via capillary action and carries the spotted compounds with it. The spots can be visualized by spraying chemical reagents onto the plane. In planar chromatography, the analyst is limited to the following variables in order to effect a separation: composition of stationary phase and mobile phase, the length of the plane and development time. Thin-layer chromatography is a nondestructive technique; the stationary phase spot containing the compound of interest can be scraped off and extracted to recover the isolated compound.

High performance liquid chromatography (HPLC) HPLC is a subdivision of LSC that uses a stationary phase that has a very small particle size in a commercially produced column that is tightly packed. The contact of the mobile phase to the stationary phase is greatly increased, resulting in superior resolution. The flow of mobile phase through such a tightly packed column is highly restricted, necessitating a quality pump capable of sustaining a high inlet pressure. HPLC can be run in either normal phase (stationary phase more polar than mobile phase) or reverse phase (stationary phase less polar than mobile phase). Since no heat is used, HPLC is ideal for the separation of volatile compounds such as high explosives and some drug samples. Advances in the last decade have allowed the outlet of an HPLC to be mated to the inlet of spectroscopic instruments such as ultraviolet, mass or infrared spectrometers, allowing automated separation and identification of complex mixtures. In

addition to the variables listed under planar chromatography, column flow rates and temperature can be varied in order to effect a separation. The ability to alter the mobile phase composition during a run is also useful in effecting separations. HPLC is a nondestructive technique and after elution the sample can be recovered by collection of the mobile phase.

Gas chromatography (GC) Probably the most widely used separation technique, gas chromatography is the subdivision of chromatography where the mobile phase is a gas (known as the 'carrier gas'). The stationary phase is either a solid adsorbent (GSC) or, more commonly, a liquid coated on the walls of the column (GLC). The separation of compounds on a GLC column is based on the relative solubility of the sample components in the stationary phase. Samples run on a GC must be first converted into a gas at the temperature of the column. This may be accomplished by dissolving the sample in a solvent and injecting the sample with a syringe into a heated (~300°C) injector port at the inlet of the column. Samples which cannot be dissolved (paint, fibers, plastics) can be converted into the vapor phase with the use of a pyrolysis unit (~800°C) at the inlet of the column. Once introduced to the column, the sample is carried through the column by the carrier gas. The column is heated in an oven to prevent precipitation of the sample and to finely control the elution of compounds from the column. Heating the column increases the solubility of the mobile phase in the stationary phase, which as a result decreases the retention time of compounds on the column. The temperature of the column can be programmed from low to high temperatures making separations highly efficient. Exiting the column, the separated substances are detected instrumentally, most commonly by a flame-ionization detector (GC-FID) or a mass spectrometer (GC-MS). The use of a mass spectrometer is most advantageous since it may allow the analyst to identify compounds as they are eluted. GC is routinely used in the forensic laboratory for the analysis of flammable compounds, toxicological samples, drugs, paint and other organic compounds (**Fig. 3**). Mobile phase flow and column

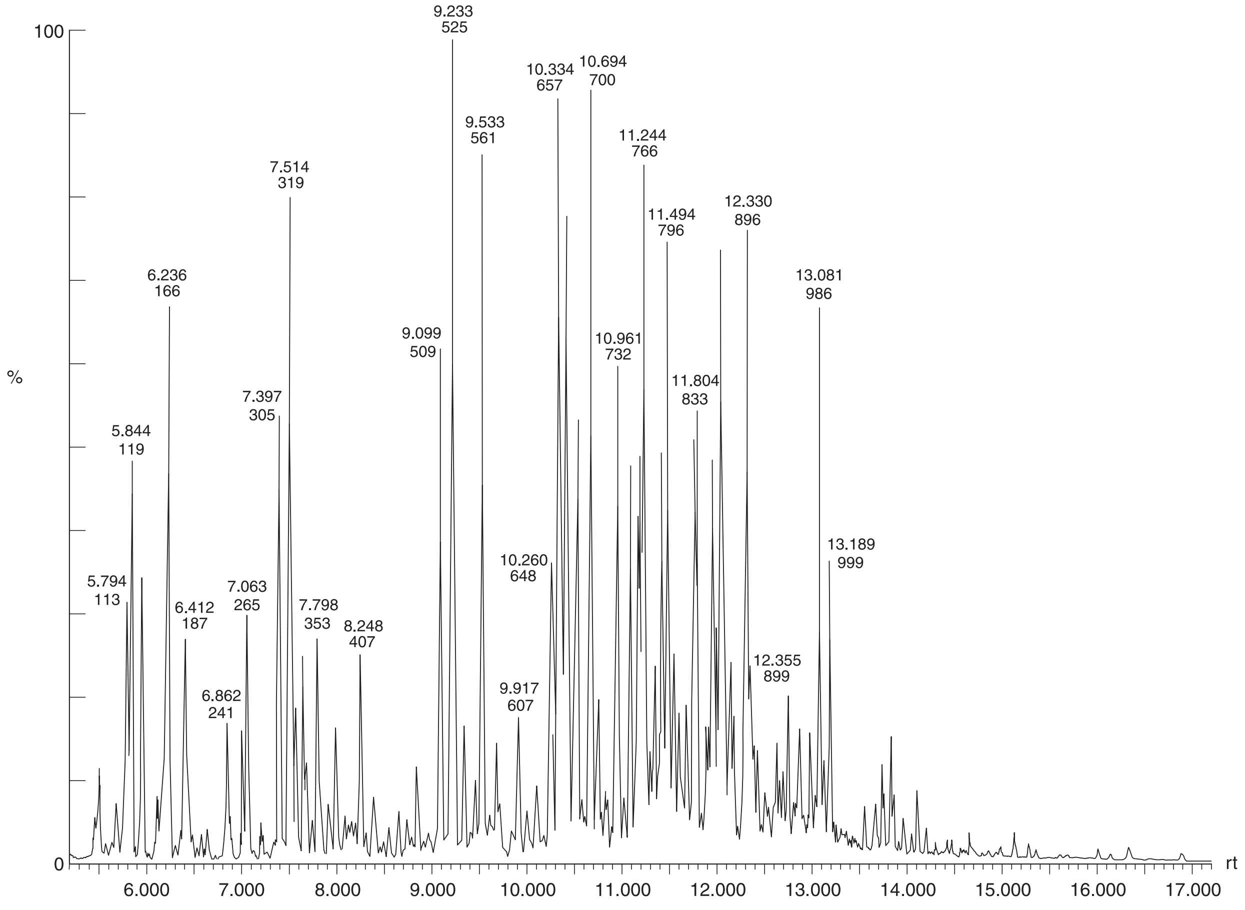


Figure 3 Gas chromatogram of unleaded gasoline. (Courtesy T. Ernst, Michigan State Police.)

temperature, which can be programmed to change during a run, are common variables used to effect a GC separation. GC is considered a destructive technique as there is no practical way to recover samples once in the vapor phase.

Electrochemical Techniques

Although these techniques bear a resemblance in theory and function to chromatographic techniques, they cannot be classified as such, as the sample compounds are not partitioned between a stationary phase and a mobile phase.

Electrophoresis

This technique is a planar technique, which uses an electric potential to separate large ions according to their mass charge ratios. The sample is spotted on a gel plate (support) which is permeated with an electrolyte solution. An electric potential is applied across the support for a fixed period of time, causing positive ions to be attracted toward the cathode and negative ions to be drawn toward the anode. The rate of migration of the ions across the support increases as the charge on the ion increases, but decreases with increasing size or mass of the ion. The separation is based on the ratio of charge to mass of the components. Changing the voltage potential across the support controls the speed of migration. The separated components can be visualized by the use of visualizing reagents. Planar electrophoresis is primarily used in forensic work in the separation of proteins present in biological fluids, including DNA.

Isoelectric focusing

The addition of a series of ampholytes to the gel results in enhanced separation of closely related ions. Ampholytes are usually polyaminopolysulfonic acids, which serve as pH buffers. When a voltage is applied, each ampholyte migrates into the gel a different distance creating a continuous pH gradient across the gel. The pH of the gel increases from the anode to the cathode. Substances migrate through the gel until they reach the pH region at which the protein exists as a neutral species. At this pH (the isoelectric point or pI), the migration ceases, since the neutral species has no attraction to the charged electrode.

Capillary electrophoresis

In the past decade, there has been a tremendous advance in the application and use of capillary electrophoresis (CE), which involves an electrophoretic separation in a narrow-bore fused silica capillary tube. The capillaries used in CE generally are about 30–50 cm in length and have an internal bore of 0.01–0.8 mm. The capillary is filled with buffer and each end is also immersed in buffer. The sample is injected at one end and an electric potential applied across the capillary, driving the components into discrete bands. At the output end of the capillary, analytes are detected and quantified. The resolution of this technique is significantly greater than planar electrophoresis and also has the added advantage of reduced run times. Capillary electrophoresis has gained a foothold in DNA laboratories and is capturing a large share of the HPLC market in forensic drug laboratories.

Ion mobility spectrometry (IMS)

Instruments, such as the Barringer Ionscan, are in their relative infancy but have become increasingly popular due to their portability, sensitivity and speed of analysis. IMS has been used as a forensic screening technique in the areas of drug and explosive residue. In this technique, samples are swabbed with a filter paper and inserted into a heated sampling device, which serves to ionize the sample. The ions are released into a separation region that is under the influence of an electric field. The ions move through this region at a rate proportional to their mass and against the flow of a gas. Negatively charged ions move quickly through the field toward the cathode and their time of flight recorded. This time of flight can be compared to the time of flight of a variety of substances across the field allowing tentative identification of the substance. These instruments are commonly used in airports to rapidly screen luggage for explosives and drugs. In the last few years they have been increasingly used at the scenes of clandestine laboratory and bombing investigations and prison inspections.

See also: **Analytical Techniques**: Microscopy; Spectroscopy: Basic Principles; Presumptive Chemical Tests; Mass Spectrometry; Gas Chromatography; Spectroscopic Techniques; Hyphenated Chromatographic-Spectroscopic Techniques, **Fire Investigation**: Types of Fire; Chemistry of Fire; Evidence Recovery at the Fire-scene; Laboratory.

Further Reading

Allen A and Cooper D (1979) Structural elucidation of low molecular weight primary and secondary amines via phenylisothiocyanate derivatives. *Microgram* **XII** (2):24–52.

Braun RD (1982) *Introduction to Chemical Analysis*. New York: McGraw-Hill.

Chamot EM and Mason CW (1940) *Handbook of Chemical Microscopy*, vol. II. New York: Wiley.

Frenkel M, Tsaroom S, Aizenshtat Z, Kraus S and Daphna

D (1984) Enhanced sensitivity in analysis of arson residues: an adsorption-tube gas chromatograph method. *Journal of Forensic Science* **29**:723–731.
Hopen TJ and Kilbourn JH (1985) Characterization and identification of water soluble explosives. *The Microscope* **33**:1–22.
Karpas Z (1989) Forensic science applications of ion mobility spectrometry. *Forensic Science Review* **1**:103–119.
Khaedi M (1998) *High Performance Capillary Electrophoresis–Theory, Techniques and Applications*. Somerset, NJ: John Wiley.
Kirkbride KP, Klass G and Pigou PE (1998) Application of solid-phase microextraction to the recovery of organic explosives. *Journal of Forensic Science* **43**:76–81.
McVicar MJ and Graves WJ (1997) The forensic comparison of soils by automated scanning electron microscopy. *Journal of the Canadian Society of Forensic Science* **30**:241–261.
Moffat AC, Jackson JV, Moss MS and Widdop B (1986) *Clarke's Isolation and Identification of Drugs in Pharmaceuticals, Body Fluids, and Post-mortem Material*, 2nd edn. London: The Pharmaceutical Press.
Oulton SR and Skinner HF (1998) Identification of common inorganic acids encountered at clandestine laboratories. *Microgram* **XXXI** (10):277–287.
Ren Q and Bertsch W (1999) A comprehensive sample preparation scheme for accelerants in suspect arson cases. *Journal of Forensic Science* **44**:504–515.
Souter RW (1985) *Chromatographic Separations of Stereoisomers*. Boco Raton, FL: CRC Press.

Spectroscopic Techniques

K P Kirkbride, Forensic Science Centre, Adelaide, South Australia

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0790

Introduction

The techniques described below are those that are most commonly encountered in forensic laboratories. For fundamental details relating to matter–radiation interactions, instrumentation, data acquisition, and data interpretation readers are directed to the reference works listed under Further Reading.

Infrared Spectroscopy

The key activity taking place during infrared spectroscopy is interaction of infrared radiation with bonds between atoms in the specimen. Absorption of radiation causes the bonds to attain a vibrational state of higher energy. As the vibrational state of a bond depends on many parameters, such as its strength, and the relative weights of the atoms at its termini, different bonds require different stimulation and therefore absorb radiation at different frequencies throughout the infrared range. Spectroscopy in the mid infrared region (4000–500 cm^{-1}) is commonly employed in forensic science. A single infrared spectrum can, potentially, provide a wealth of information relating to a specimen.

Infrared spectroscopy is unsurpassed as a universal technique for the identification of the class of compounds present within a specimen. For example, it is very easy to identify fibers on a generic basis such as acrylics, nylons and polyesters, or paints on a generic basis such as alkyds, acrylics, urethanes, nitrocellulose, etc. With the advent of computerized spectral databases and chemometrics it is now possible to extend the technique to allow subgeneric discrimination.

However, in a few circumstances infrared spectroscopy might not allow easy discrimination within a class, particularly if mixtures of members of the class are present. **Figure 1** gives the spectral data for three common phthalate ester plasticizers. Although dimethyl phthalate is easily distinguished from either diisooctyl phthalate or dihexyl phthalate, it is very difficult to distinguish the latter closely related esters from each other. Consider, then, two polymer specimens, one composed of polyvinyl chloride plasticized with diisooctyl phthalate, and another composed of polyvinyl chloride plasticized to the same extent with a mixture of diisooctyl phthalate and dihexyl phthalate. Infrared spectroscopy would not be capable of

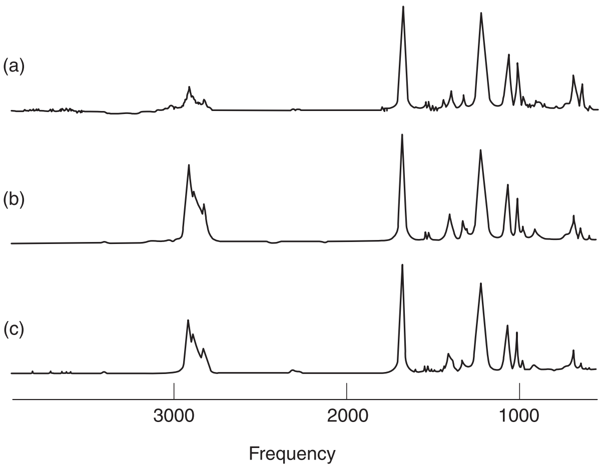


Figure 1 Infrared spectral data from three phthalate ester plasticizers. (a) Dimethyl phthalate; (b) diisooctylphthalate; (c) dihexylphthalate.

distinguishing these two specimens. Another good example of the limitation of the technique is differentiation of paints within the alkyd class.

The limit of detection, using infrared spectroscopy, of a substance in the presence of another is about 5%. Infrared spectroscopy is therefore not a good technique for the detection or identification of trace copolymers, residual solvents, low-level pigments, or some additives in polymer samples (e.g. fibers or paints), nor is it of any use for the detection or identification of manufacturing impurities, or trace contaminants in illicit drugs. Above this limit of detection, however, infrared spectroscopy can be a reliable qualitative tool, and even semiquantitative analysis can be performed. For example, by using a set of calibration data comprising various known mixtures of heroin in sucrose, relatively accurate quantitative analysis of unknown mixtures of heroin in sucrose can be performed. Unlike chromatographic techniques (such as liquid chromatography or gas chromatography) that are somewhat independent of matrix identity, matrix contribution must be accounted for in quantitative infrared spectroscopy. It would not be possible, for example, to estimate the level of heroin in glucose using calibration data obtained from mixtures of heroin in sucrose. Specific calibration sets must be established for each type of analysis.

Infrared spectroscopy can be used to elicit structural information that might be difficult or impossible to arrive at using other techniques. One area of great utility is the characterization of stereoisomers. For example, the diastereoisomers ephedrine and pseudoephedrine are not distinguishable by mass spectrometry, and depending on the stationary phase used, might exhibit identical retentions in a gas chromatograph. However, infrared spectral data for these two compounds are easily distinguishable. In a similar fashion infrared spectroscopy is very efficient as a means of discrimination between positional isomers of aromatic substances (e.g. the various isomers of dimethoxyamphetamine, or trimethyl benzene) and between other stereoisomers (e.g. the isomeric group cocaine, allococaine, pseudococaine, and pseudoallococaine; or the group butyl nitrite, isobutyl nitrite, and secbutyl nitrite, or the isoheroins and heroin).

It is often of some importance to ascertain whether a drug is present as free base or a salt. The N–H stretch in an amine salt is found between 2500 and 3000 cm^{-1}, whereas the N–H stretch in its free base is found well above 3000 cm^{-1}. As a consequence, spectral properties for drug free bases and their salts are very different in the 2500–3500 cm^{-1} region.

Infrared spectroscopy is a technique that can provide information relating to the total composition of the specimen, unlike chromatographic techniques, which usually require an extraction step and have restrictions due to volatility, solubility, etc. Such techniques therefore can only provide partial information as to the specimen. In the examination of paint for example, infrared spectroscopy can provide information as to the polymeric binder as well as the inorganic extenders and pigments. **Figure 2** shows the spectrum of a paint specimen, and spectra for kaolin and silica. Spectral features due to those minerals as well as the polymer are evident in the spectrum of the paint. Inorganic halides are virtually transparent to infrared radiation, therefore the presence of, for example, common salt or potassium bromide in a specimen could not be deduced. Metals and some of their oxides also do not give useful data.

The very feature that can make infrared spectroscopy useful (i.e. sensitivity to a wide range of substances) might also be a disadvantage. **Figure 3** shows

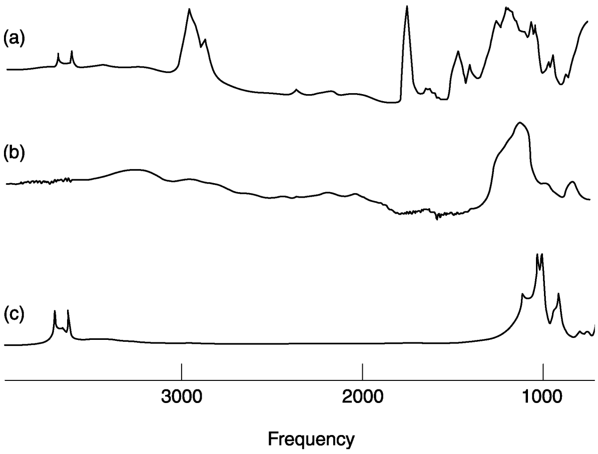


Figure 2 Infrared spectroscopy of paint films allows identification of inorganic compounds as well as the polymeric binder. In this example neither the binder nor the inorganic substances dominate the spectrum. (a) Paint film; (b) infrared spectrum for silica; (c) infrared spectrum for kaolin.

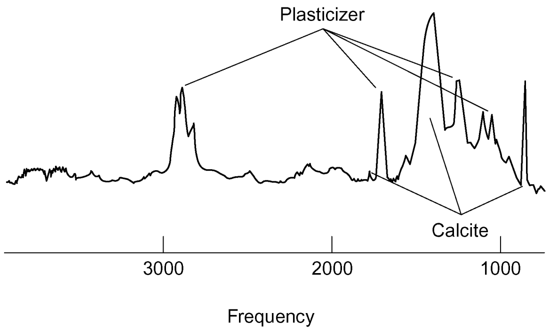


Figure 3 Infrared spectrum recorded from a small piece of plastic electrical insulation. Peaks indicated are due to plasticizer or calcite. In this spectrum the inorganic fillers dominate, therefore the polymeric substance (polyvinylchloride) is very difficult to identify.

the spectrum obtained from a sample of colored matter (apparently electrical wire insulation) taken from a pair of wire cutters suspected of being used to disable a burglar alarm. Spectral features due to the polymer (PVC) are almost nonexistent, the majority of peaks are due to plasticizer (a phthalate ester, or a mixture of them) and calcite. As a means of comparison between the recovered chips of plastic and insulation taken from wire in the burglar alarm, infrared spectroscopy is particularly poor. Spectral features due to a very important component of the material, the polymer, are masked by peaks due to additives.

Undercoats and primers can be very high in inorganic solids such as calcite, talc, kaolin etc. It is often the case that the presence of these substances precludes infrared spectroscopy as a means of identifying the organic binder, or makes it very difficult. This does not mean that infrared is a poor technique for discrimination of highly filled paints, but it does mean that any discrimination is likely to be based primarily on fillers, and not on polymeric composition. Interpretation of results should, therefore, bear this in mind, or another technique to identify the polymer composition should be carried out (such as pyrolysis–gas chromatography–mass spectrometry).

Infrared spectroscopy can be carried out on microscopic specimens (down to about 10 μm square) using an infrared microscope. This instrument is simply a device which condenses the infrared beam from its standard size (usually about 10 mm) to much smaller dimensions (about 0.18 mm). This task is accomplished with a microscope, which is also used to allow high magnification viewing of the specimen and accurate positioning of it in the beam. The technique is referred to as infrared microscopy or, more accurately, infrared microspectroscopy.

Infrared microspectroscopy is the usual technique employed in the analysis of small samples of paint (e.g. motor vehicle accidents, breaking and entering), single textile fibres (evidence of contact), particulate material (explosives residues such as ammonium nitrate, smokeless powder, etc.), and concentrated extracts of low dosage drugs (e.g. an extract from an LSD (lysergic acid diethylamide) dose evaporated to a microscopic spot). It is possible to record spectra from a few nanograms of material using an infrared microscope.

In practice, infrared microspectroscopy differs little from infrared spectroscopy. Sample preparation is a little more tedious, as manipulations must be carried out with the aid of a microscope. In most instances steps must be taken to reduce the thickness of the specimen to about 10–30 μm, otherwise bad artifacts will be evident in the spectrum. This can be accomplished by using a sharp blade, flattening the specimen between diamond plates, rolling the specimen with a small metal roller, microtoming the specimen, etc. Even though it is possible to see very small objects (less than 10 μm) with the infrared microscope, diffraction of the infrared beam inhibits the acquisition of high quality spectra. A description of diffraction and its effects on microspectroscopy is beyond the scope of this text, the reader is referred to the bibliography for useful references.

With infrared microspectroscopy, as with any other microspectroscopic technique, the homogeneity of the specimen is an important issue. For example, if single textile fibers are contaminated with blood, adhesive from tape lifts, or mountant (if the fiber has been examined previously with the aid of a compound microscope), then the contaminant is likely to give rise to unwanted peaks in the spectral data. In the case of some paint samples the granule size of some inorganic fillers, such as calcite, can be relatively quite large. Therefore, on the microscopic scale some paints are quite heterogeneous. Effects of heterogeneity can be noticed if one is forced to analyze a very small specimen of highly filled paint (e.g. architectural undercoats), or if a very small probe beam is used. Very small specimens might not exactly represent the composition of the macroscopic film from whence it came, by chance the specimen might contain more or less calcite, for example. Similarly, if a relatively large specimen is taken for analysis, but only a very small portion of it is probed by the infrared beam, the spectrum produced will depend on exactly where the microbeam strikes the specimen. In some locations the beam might encounter mostly calcite, whereas in other locations the beam might encounter mostly polymeric binder. In infrared microspectroscopy, therefore, one should take steps to insure that an accurate representation of the specimen has been achieved. If it has been possible to recover a relatively large specimen, then as much of it as possible should be analyzed. On the other hand, if only tiny specimens can be recovered it is wise to analyze a few of them (separately) in order to fully represent the film from whence they came. A single, minute paint chip presents a problem one must be aware that it might not be representative of its source.

The foregoing has described infrared spectroscopy transmission–absorption techniques, that is, infrared radiation is allowed to pass through the specimen, and absorption of radiation is measured. There are two other techniques in widespread use in forensic laboratories; diffuse reflectance spectroscopy (DRIFTS) and internal reflectance spectroscopy (IRS, sometimes referred to as attenuated total reflectance (ATR) spectroscopy).

In IRS the infrared beam is caused to reflect inside a

special crystal. At the point of reflection the beam breaks through the crystal surface slightly (the order of about 1 μm) to form what is referred to as an evanescent wave. The specimen is placed on the crystal at the point of the evanescent wave and any absorption of the wave by the specimen is recorded as the IRS spectrum (**Fig. 4**). There are some important considerations with IRS. First, the magnitude of the evanescent wave is dependent on the wavelength of the radiation; long wavelength infrared generates a bigger evanescent wave, short wavelength infrared a smaller wave. The outcome is that IRS spectra, compared to standard transmission spectra, show increased absorbances of peaks at long wavelength. Second, the penetration of the evanescent wave into the specimen is small because the wave itself is small, therefore IRS is a surface analytical technique. IRS will overrepresent surface contamination or surface modification, if the specimen is laminated or graded in some way then IRS will provide data relating to the surface only. For comparative work therefore, IRS data should not be compared with transmission data. Furthermore, the properties of IRS should be borne in mind when examining sheath-and-core bicomponent fibers, and identifying polymers with surface films, surface contamination or surface oxidation.

Diffuse reflectance is a special type of reflection from the specimen; the origin of the phenomenon is not of great relevance here. The important issue is that during the process of diffuse reflection the infrared beam interacts with the specimen and some radiation is absorbed. It is this absorbed radiation that is measured and plotted, versus frequency, as the DRIFTS spectrum.

The main virtue of DRIFTS is the ease with which samples can be prepared. All that is required for powdered specimens is to insure that the particle size is reasonably small, the spectrum can be acquired directly from the specimen raked flat in a small cup. Solutions of drugs, for example extracts in dichloromethane, can be applied to ground potassium bromide in the sample cup and their spectral data collected after the solvent has evaporated. Large specimens of polymer can be abraded using silicon carbide paper, and the spectrum of the dust thus produced can be acquired directly from the surface of the paper. DRIFTS is representative of the surface of a specimen, therefore, depending on the exact method of specimen preparation, the precautions listed above for IRS might need to be borne in mind.

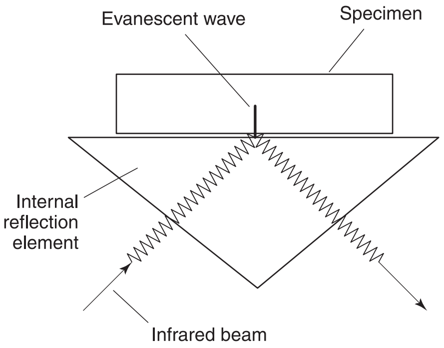


Figure 4 Principle of internal reflection spectroscopy. The infrared beam enters a special crystal and is internally reflected off one of its surfaces. At the point of internal reflection an evanescent wave penetrates the crystal surface into the specimen, which is in very close contact with the surface. The beam, which has suffered some absorption via the evanescent wave then travels on to the detector.

Visible and Ultraviolet Spectroscopy

Visible (Vis) spectroscopy and ultraviolet (UV) spectroscopy are usually referred to as a single technique (UV–Vis). This is because the two types of radiation are adjacent in the electromagnetic spectrum, the molecular transitions that give rise to the spectroscopic phenomena are of the same nature, and usually a single instrument is capable of acquiring both types of spectral data. The key activity taking place during this technique is interaction of visible light or ultraviolet radiation with electrons within molecules in the specimen. Absorption of radiation causes molecules to attain an excited state of higher potential energy. The energy of the excited state depends on the types of bonds and atoms present in the specimen, therefore different electrons in different substances require different stimulation, and therefore absorb different frequencies of radiation. This probe, however, is not as sensitive to subtle changes in molecular structure as say, infrared spectroscopy or nuclear magnetic resonance (NMR), therefore UV–Vis spectroscopy is a technique with relatively limited discriminating power. As a consequence, UV–Vis spectroscopy does not feature prominently in forensic laboratories as an identification technique.

A major use is screening powders for the presence of illicit drugs. In this application UV–Vis is quite successful because the technique has a very low limit of detection, powders are usually simple mixtures, and illicit drugs give good signals.

UV–Vis is a very useful, if somewhat limited, quantification technique, again particularly for illicit drugs. Of all the spectroscopic techniques it is easiest with UV–Vis to make use of Beer's law, which states that the absorbance of radiation at a given frequency is proportional to both the distance the radiation travels through a specimen and the concentration of the species that absorbs the radiation.

$$A = \varepsilon \times C \times L$$

where A = absorbance; ε = extinction coefficient; C = concentration of the absorbing species and L = path length of radiation through the specimen.

The constant of proportionality, ε, in Beer's law is called the extinction coefficient. It is a measure of how well a compound absorbs radiation at a given wavelength; it is as characteristic of that compound as other physical constants such as melting point, boiling point or molecular weight. The chemical literature abounds with extinction coefficients for many substances, including pure drugs and their salts in a variety of solvents at a variety of wavelengths. Many useful solvents such as water, methanol, ethanol, and acetonitrile are virtually transparent throughout the visible and ultraviolet spectrum. It is, therefore, very easy to prepare a solution of the illicit preparation to an accurate concentration in a solvent that does not interfere with the absorption spectrum. Using Beer's Law, the absorbance of the specimen, the optical path length (usually 1 cm in standard sample cells), and the extinction coefficient of the drug, the concentration of drug present in the solution of the illicit preparation can be calculated. The attraction of such an analytical strategy is that it is very inexpensive, rapid, and simple. Furthermore, if the UV–Vis spectrometer has been properly calibrated using absorbance and wavelength references, a pure reference drug standard is not absolutely necessary in order to quantify the drug because extinction coefficients are universal constants.

This facility for accurate, standardless quantification is unique to UV–Vis spectroscopy; chromatographic techniques (e.g. GLC, HPLC, CZE) require a reference substance in order to calibrate the response with respect to amount, and for other spectroscopic techniques (e.g. infrared spectroscopy) extinction coefficients are not available. UV–Vis spectroscopy is not generally applied to the quantitative assay of illicit drug preparations, however, because the technique in general has severe limitations for the analysis of uncharacterized mixtures. However, when a mixture is well characterized and there is some degree of separation between the spectral peaks arising from the compounds in the mixture, UV–Vis spectroscopy can be useful.

Another example is the estimation of carboxyhaemoglobin in human blood. The spectral characteristics (i.e. absorption maxima and extinction coefficients) for carboxyhaemoglobin and interfering blood pigments (e.g. oxyhaemoglobin, methaemoglobin) are well known. For a blood specimen of unknown carboxyhaemoglobin saturation it is a simple matter to deconvolute mathematically the spectral contributions due to interfering blood pigments from the bands due to carboxyhaemoglobin.

A major application of visible spectroscopy is the objective color comparison of evidential material such as fibers or paint. The human eye and brain interpret the visible spectrum of light reflected from an object as its color. However, the human eye is a low-resolution transducer, it can be fooled into interpreting two closely related colors as identical. **Figure 5** shows visible spectra acquired from two different turquoise wool fibers. Even though the spectra from the two fibers are noticeably different, the eye perceives the two fibers to be identical in color. In the modern forensic laboratory, therefore, visible spectroscopy is invaluable as a means of confirming whether two objects are of the same color, or refuting the observation.

Conventional UV–Vis spectrometers cannot acquire spectral data from microscopic samples of paint, nor from short pieces of single textile fibers. The instrument used is referred to as a visible microspectrometer, it functions in the same way as the infrared microscope described above. Unlike the infrared microscope, visible microspectrometers allow analysis and visualization of objects less that 10 μm in size. This is possible because visible light suffers less diffraction than infrared radiation. Visible spectroscopy should be considered a color analysis technique, not a pigment analysis technique. This is because visible spectroscopy is not always capable of resolving different compounds (or different blends of compounds) that have the same color. Reliable pigment analysis is achieved only when visible spectroscopy is used in conjunction with other techniques. For example, thin layer chromatography (TLC) could be attempted on pigments extracted from fibers, or scanning electron microscopy–energy dispersive x-ray microanalysis could be used to characterize pigments in paint.

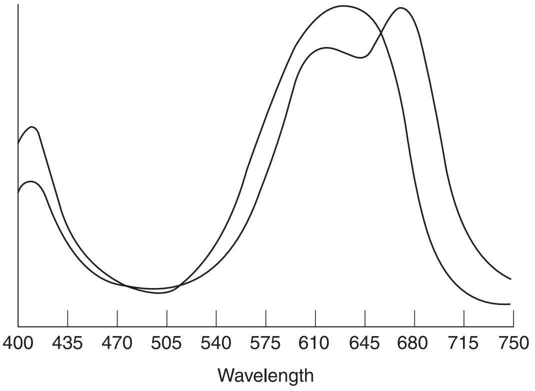


Figure 5 Visible microspectra of two different, metameric turquoise wool fibers. Although these fibers are of an identical color to the human eye, they are easily distinguished using visible microspectroscopy.

Pigments absorb radiation in the UV region as well as in the visible region, and paints designed for outdoor application can contain compounds that are active in the UV. Therefore, compared to visible spectroscopy, UV–Vis spectroscopy can be expected to offer enhanced discrimination. However, the standard optics present in visible microspectrometers are usually made of glass, and these strongly attenuate UV radiation. It is possible to purchase microspectrometers equipped with quartz optics, which allow the spectrometer to cover the UV and visible range. When faced with the high cost of quartz optics however, forensic laboratories tend to rely on visible microspectroscopy and enhance pigment discrimination with a simple, inexpensive technique such as TLC.

Nuclear Magnetic Resonance Spectroscopy

In NMR spectroscopy the absorption of radiation (radio waves) by the specimen causes the nucleii of atoms in it to attain excited states of spin. The exact energy of the excited spin states depends very markedly on the molecular environment the atoms are in. Nuclei in different environments therefore absorb different frequencies within the radio wave region. There is such a sensitive correlation between environment and the signal associated with the nuclei that NMR is capable of yielding an enormous amount of information regarding molecular structure, and therefore unsurpassed specimen discrimination.

In a forensic context NMR finds its greatest application in the analysis of organic compounds. It can be described as a wide range technique, applicable to volatile, low molecular weight compounds as well as involatile, relatively high-molecular-weight substances. Substances of high polarity or low polarity can be analyzed, and it is usual to conduct the analysis at near to room temperature, therefore thermally labile substances can be analyzed. A few factors have, however, conspired to minimize the impact of NMR spectroscopy on forensic science. First, instrumentation is expensive to purchase and, in the case of instruments equipped with superconducting magnets, expensive to maintain. Second, analytical mixtures must be relatively simple as commercial instruments with interfaces to chromatographic techniques are not available. Finally, the technique has a relatively high limit of detection, and unfortunately that limit is inversely proportional to the cost of the instrument.

Analysis is usually conducted in one of two modes, proton mode which detects signals from the nuclei of hydrogen atoms, and carbon-13 mode which detects signals from natural abundance ^{13}C nuclei. In carbon-13 mode it is usual to see a signal for every nonequivalent carbon atom within the specimen. Therefore, if the specimen is a pure sample heroin, for example, 21 distinct signals at characteristic frequencies are observed. This is an extremely high level of discrimination. In proton mode, the situation is analogous. Let us now consider the case where the specimen is an equimolar mixture of acetyl codeine, O^6-acetylmorphine, and heroin. In this case the carbon-13 spectrum would contain 60 signals, some resolved, some not. If the task was a target analysis of the specimen for the presence of heroin, then it would be difficult but successful. If the task was identification of an unknown specimen, it would be very difficult indeed. Therefore one of the strengths of NMR, the provision of enormous structural detail, becomes a limitation in the case of complex mixtures. NMR is therefore not applicable to the identification of hydrocarbon fuels, for example. In polymers and oligomers many atoms have very similar environments, therefore many signals overlap. NMR therefore is not valuable for the forensic discrimination of paints or fibers.

NMR is not a technique that can be applied to trace analysis, for example detection of explosives in bombing debris, or drugs in body fluids.

NMR can be of great value for target analysis of illicit drugs because the analytes are usually present in high concentration in a simple mixture. Specimen preparation is very simple; the substance can be dissolved in a suitable solvent (heavy water, deuterochloroform, deuterated methanol) or basic drugs can be extracted from alkaline aqueous solution using a suitable solvent (deuterochloroform, carbon tetrachloride). If the instrument is calibrated and an internal standard is included in the protocol, then the test can be quantitative as well as qualitative. The extremely high level of discrimination gives the test a very high probative value. Due to the nature of material encountered, NMR is also very useful in the identification of precursors and intermediates found in clandestine laboratories.

One of the greatest benefits of NMR spectroscopy is the predictive structural information that can, with skill, be elucidated from the spectrum of an unknown substance. From the carbon-13 spectrum the number of distinct carbon atoms within the unknown molecule can be deduced. By sophisticated pulse techniques these can be identified as methyl, methylene, methine, or quaternary residues, and by careful examination of the position of each peak, the presence of heteroatoms and carbon–carbon multiple bonds can be inferred. This information allows a good picture of the component bits and pieces of the unknown molecule. Proton NMR allows the

connectivity of the moieties to be deduced, and a complete picture of the molecular structure built up. Using this technique it is possible to identify new designer drugs and unusual precursors or intermediates in clandestine laboratories.

X-ray Fluorescence

In this technique the specimen is irradiated with X-rays. This causes electrons associated with atoms in the specimen to attain an excited state of potential energy. From this excited state the electrons can relax, or release that potential energy, a process which results in the specimen emitting energy (fluorescence) in the form of new X-rays. The X-rays are emitted at energies that are characteristic of the atoms present in the specimen. Therefore the X-ray fluorescence (XRF) spectrum carries information as to the chemical elements present in the specimen.

XRF spectrometers can be manufactured for the analysis of microscopic specimens. Given that it is necessary in forensic science to have the capability for microanalysis this section will deal only with micro-XRF.

In this technique X-rays are delivered to the specimen via a light guide. It is possible to bring the beam into 'focus' in a spot about 10 μm in diameter. The specimen is viewed under low magnification with the aid of a video camera, which allows control over the area illuminated by the X-ray beam and therefore the region in the specimen that is analyzed.

The X-ray beam even in a micro-XRF spectrometer penetrates a long way into the specimen and interacts with a relatively large volume of the specimen. Therefore, with thin, microscopic specimens, such as small flakes of glass or thin pieces of paint, most of the X-ray beam will pass right through and fluorescence will be minimal. If the specimen is constructed of thin laminates then special sample preparation, such as the preparation of cross-sections or subsamples dissected from the entire specimen, must be undertaken unless a composite spectrum of all the laminations is required.

Given the sampling and imaging idiosyncrasies of XRF, it is useful for the elemental analysis of relatively large, easy-to-find specimens such as thickly layered paints, big chips of glass, fibers, and pieces of automotive accessories. Compared to scanning electron microscopy (SEM)–X-ray microanalysis, XRF has slightly lower limits of detection for elements in a homogeneous specimen, therefore there is the potential for a slightly higher level of discrimination in the analysis of metal alloys and glass for example. Compared to techniques such as atomic absorption spectroscopy, however, detection limits are relatively high.

Many specimens in forensic science are quite heterogeneous on the microscopic scale and useful information might arise from probing the microstructure. This facility is not especially well catered for by XRF due to the low contrast images produced by the video system, the limited range of specimen magnification achievable, and the large volume of specimen that is excited by the probe beam. Illustrations of the limitations of the technique are provided by paint comparison and firearm primer analysis.

In paint comparison, analysis of a relatively large volume of specimen using XRF indicates the elements present down to the limit of detection for the technique. Let us assume that XRF indicates the presence of Ca, S and Ba. The technique does not allow us to discriminate between paints that contain calcium sulphate and barium sulfate, or barium sulfate and calcium carbonate, or a mixture of calcium sulfate, calcium carbonate and barium carbonate. In order to resolve this dilemma what is needed is the ability to distinguish individual particles within the paint film and analyze them. However, with the video imaging system associated with XRF it is difficult to discern individual filler particles unless they are very coarse grained, and the relatively large volume excited by the probe beam means it can be difficult to restrict excitation to individual particles.

With firearm primer particles on a tape-lift the analyte is composed of minute particles (usually less than 10 μm) scattered thinly over a relatively large area. It is not possible with XRF to enhance specimen image contrast (i.e. by techniques such as backscattered electron imaging as used in SEM–X-ray microanalysis) or achieve high magnification, therefore it is a very difficult task to locate analyte particles. Furthermore, the particles are likely to be much smaller than the probe beam, and in any event the volume of specimen excited by the beam is relatively large, therefore useful fluorescent yield will be low, and contribution from the mounting medium could be large.

Compared to micro-XRF, analysis of heterogeneous specimens is likely to be more efficient using SEM–X-ray microanalysis (see below).

The application of XRF to the analysis of explosives residues further illustrates the versatility of the technique, but also a shortcoming that is common to any X-ray microanalytical technique. For example, XRF can be used to identify particles containing chlorine arising from chlorate, perchlorate or hypochlorite-containing improvised explosive devices. Unfortunately it is not possible to use XRF to distinguish these oxyanions from each other, nor from chloride, which is not characteristic of explosives residues. Similarly, in cases of black powder explosions, XRF

can be used to identify the presence of potassium and sulfur, but it is not possible to prove whether the sulfur is present as one of its oxyanions, or as elemental sulfur, or sulfide.

Molecular Fluorescence

As described above, when molecules absorb visible or ultraviolet radiation they are elevated to an excited state of high potential energy. From that state the molecule can undergo a variety of actions in order to relax. One activity is to emit radiation of lower energy than that which caused the excitation in the first place. The emitted radiation (fluorescence) consists of many frequencies of characteristic intensity. Like the UV–Vis spectrum, the emission spectrum (i.e. the plot of emitted radiation intensity versus wavelength) of a substance can be a means of charaterizing that substance.

Fluorescence has three important applications in forensic science. Perhaps the most widespread use is not in relation to spectroscopy at all, but image or specimen enhancement in the fields of ink comparison, latent fingerprint enhancement, and classification of textile fibers. Another use is in relation to chromatographic detection associated with sepaDration techniques, such as high performance liquid chromatography and capillary electrophoresis, where extremely low limits of detection can be realized. Finally, molecular fluorescence can be used as a spectroscopic technique (spectrofluorimetry).

Although spectrofluorimetry is related to UV–VIS spectroscopy, there are some significant differences. One distinguishing property of fluorescence, a feature that is both its biggest asset and its biggest drawback, is that not all molecules exhibit this behavior. In general terms, only aromatic compounds or highly conjugated systems are fluorescent, particularly if oxygen or nitrogen atoms are attached. The presence of electron-withdrawing groups (e.g. halogen atoms, nitro groups) diminishes or even completely quenches fluorescence in these substances. On the positive side therefore, techniques based on fluorescence can have very low limits of detection because spurious background signals are not likely to arise. As the excitation radiation is of a shorter wavelength that the fluorescence, it is possible to enhance detection further by filtering out interference from the exciting radiation. Spectrofluorimetry has a potential for high discrimination due to the low number of fluorescent compounds. Furthermore, the exact nature of the emission spectrum is linked to the nature of the excitation, and this has a positive effect on discrimination. For example, if two compounds have similar emission spectra, there is the potential for further discrimination if different excitation is required to produce that emission. By means of sophisticated instrumentation (a description of which is beyond the scope of this article) it is possible to scan the excitation radiation in a stepwise or continuous fashion and record emission data.

On the negative side, techniques based on fluorescence will have a very limited applicability because not all compounds fluoresce. As petroleum-based greases, oils, and concentrated fuel residues contain aromatic compounds (in particular polynuclear aromatic compounds) spectrofluorimetry can be used as a means to characterize them. Some drugs (e.g. LSD, phenothiazine, quinine and quinidine) are fluorescent, and therefore amenable to qualitative and quantitative analysis using spectrofluorimetry. Some metal atoms can also fluoresce, this finds some application in glass analysis.

Raman Spectroscopy

In terms of analytical results, but not mechanism, Raman spectroscopy very closely resembles infrared spectroscopy. The Raman signal arises through scattering of a probe beam of (usually) visible light. Scattering is a phenomenon whereby incident radiation interacts with molecules in the specimen, thereby raising them to an excited state. From this excited state molecules relax and emit radiation. During this process some molecules might retain some energy, therefore the scattered radiation will be of lower energy than the probe beam, or some molecules might release more energy than they received, in which case the scattered radiation is of higher energy than the probe beam. The energy kept by the specimen results in excitation of its vibrational states, whereas energy released from the specimen comes from relaxation of its vibrational states. The scattered radiation therefore contains information relating to the vibrational states of the molecules in the specimen. The spectrometer usually detects only the scattered radiation of lower energy and generates a plot of frequency of scattered radiation versus intensity. This plot is the Raman spectrum, and the peaks present are due to vibrational excitation; it is for this reason that Raman spectroscopy is referred to as the complement of infrared spectroscopy.

As the probe beam comes from a laser operating close to the visible region it is not absorbed strongly by water or glass, as is the case with infrared radiation. Therefore, one advantage of Raman over infrared spectroscopy is that specimens can easily be analyzed as solutions in water or as films on a glass microscope slide. In infrared spectroscopy specimens must be supported on halide crystals, which are soluble in

water and easily ruined, or on insoluble but expensive crystals such as diamond, zinc selenide etc.

As an analytical technique Raman spectroscopy has a lot to offer the forensic scientist, but it is not widely used. Although the discrimination of the technique is as good as infrared spectroscopy, it does have some drawbacks compared to infrared. First, infrared spectroscopy is a well-established, mainstream technique requiring relatively inexpensive instrumentation. Second, arising from the foregoing, there is a wealth of infrared spectroscopic data compiled for many substances; e.g. fibers, drugs, minerals, paint etc. Third, the incident radiation can induce fluorescence, which can swamp the weak Raman signal. For these reasons Raman spectroscopy might best be described as a technique of emerging importance for forensic science. One area in which Raman spectroscopy promises strong potential, compared to infrared spectroscopy, is microspectroscopy. As visible light is used as the probe beam in Raman microspectroscopy, there is less beam diffraction. Therefore much smaller specimens can be analyzed using Raman microspectroscopy (~1 μm as opposed to ~10 μm) and spatial resolution is much better. This means that if two different substances are present side by side, such as adjacent layers of paint or the two components of a bi-component fiber, it is much easier to achieve a spectrum of one of the substances free of spectral interference from the other. As Raman spectroscopy is not a transmission technique, there is no requirement to make the specimen transmit radiation, as is the case in infrared microspectroscopy, therefore specimen preparation is very simple. Currently Raman microspectrometers operate with the probe beam impinging on the surface of the specimen, therefore the technique can overrepresent the surface chemistry of the specimen. Future developments in confocal Raman microspectroscopy could allow accurate control over the probe beam so it can be brought into focus in a very small volume within the specimen, not just on its surface. This will allow evaluation of the entire specimen, and might be a useful way of avoiding excess fluorescence arising from surface contamination, textile brighteners, etc.

Atomic Spectroscopy: Atomic Emission Spectroscopy, Atomic Absorption Spectroscopy and Related Techniques

In atomic spectroscopy atoms within the specimen are elevated to an excited state by extreme heat as provided by a flame or a plasma torch. From this excited state atoms can emit radiation (atomic emission spectroscopy), or absorb radiation from a probe beam (atomic absorption spectroscopy (AAS)). The spectrometer detects absorbed or emitted radiation that is characteristic of the elements present in the specimen. In a related technique, excited atoms produced by a plasma torch are passed on to a mass spectrometer which differentiates the elements present on the basis of their atomic weight. This technique is called inductively coupled plasma–mass spectrometry (ICP–MS), and strictly speaking it is not a spectroscopic technique. As it is closely related to the other techniques described in this subsection it will be discussed along with them.

Atomic spectroscopy finds some application in forensic science. In classical experiments, the analyte is presented to the instrument as a solution in a relatively large volume of solvent. As a means of solution analysis atomic spectroscopy offers the forensic scientist unsurpassed limits of detection, accuracy and precision for elemental analysis. This does not imply, however, that it provides the most effective means by which a forensic scientist can conduct elemental analysis. First, classical atomic spectroscopy is destructive, the sample presented for analysis is usually treated with a very strong acid to form a solution, and then irreversibly aspirated into the instrument. Second, because the sample is homogenized by dissolution, atomic spectroscopy cannot yield any information as to spatial distribution, or compounds present in the specimen. For example, a sample of paint might be found to contain Ca and S. Although this strongly suggests that the paint contains calcium sulfate, one cannot rule out the possibility that calcium carbonate and a sulfur-containing compound are present, or that the paint might contain granules of calcium carbonate and calcium sulfate. Third, any contaminant associated with the specimen will be digested along with it, and will contribute to the results. Fourth, although atomic spectroscopic techniques do have very low limits of detection, they are often not low enough to detect trace elements in trace evidence. This is because the specimen must be made into a solution of relatively large volume (usually 0.5–5 ml). As a consequence trace elements in, for example, small chips of glass or paint yield very dilute solutions. Finally, some techniques, such as flame atomic absorption spectroscopy, only allow sequential analysis of target elements; one analytical test provides data with respect to only one element. As it is not possible to screen a specimen for many elements in the one test the analysis is not particularly efficient, especially with regard to specimen consumption.

Paradoxically, given the very low limits of detection for these techniques, they are of greatest use in the analysis of a relatively large specimens, and given that the technique is destructive, specimens must be big enough to allow subsampling. Such specimens

could be human tissue for toxicological analysis, and milligram-size pieces of glass, paint and metals.

Another strong application of atomic spectroscopy is the analysis of illicit drug powder samples. The low limits of detection that can be achieved allow many trace elements to be detected in heroin, for example. It is possible to identify the source country of the drug on the basis of the suite of elements it contains.

Some of the major shortcomings of ICP analysis can be rectified by the use of a laser ablation source. In this technique a laser beam is use to vaporize very small quantities of the specimen which are then swept into the instrument. It is possible to allow the laser beam to dwell on the specimen for some time before analysis, thereby effectively removing any surface contamination. As the laser beam can be focused to a small spot size, it is possible to sample and analyze discrete regions within the specimen. This allows some identification of the spatial distribution of compounds within the specimen. Finally, the laser ablates only a tiny amount of material, therefore the specimen is left intact for further analysis.

Spectroscopy-related Techniques

Scanning electron microscopy–X-ray microanalysis (SEM–EDX)

Scanning electron microscopy–X-ray microanalysis strictly speaking is not a spectroscopic technique, however, with respect to operation and application it is very similar to a spectroscopic technique.

The basis of the instrument is an electron microscope, which in order to generate an image of the specimen, bombards it with a beam of electrons. Many phenomena originate from the collision between electrons and the atoms in the specimen, but the most important as far as this article is concerned are scattering of the electron beam, and emission of X-rays. X-rays are emitted when the atoms within the specimen capture energy from the electron beam and attain an excited state. From this state the atoms can relax and emit X-rays. The key feature of the technique is that X-ray emission is not random; different atomic elements within the specimen emit rays with characteristic energy. Therefore the X-ray emission contains information as to the atoms present within the specimen, in this regard the technique resembles XRF.

In order to produce a spectrum from the emitted X-rays they are processed in one of two ways. One way is to pass the X-rays through a special crystal that acts like a prism and disperses the radiation in space. This technique is referred to as wavelength dispersive spectroscopy (WDS). Another technique involves collecting the X-ray emissions from the specimen and sorting them on the basis of their energy. The intensity of X-ray emission is then plotted as a function of energy; this approach is called energy dispersive X-ray microanalysis (EDX).

Although wavelength dispersive X-ray spectrometers offer superior performance in terms of resolution and limit of detection, EDX spectrometers are more common in forensic laboratories due to their relatively low cost. Here, only EDX is referred to, although the comments made are equally applicable to WDS.

Combination of EDX with scanning electron microscopy makes a versatile and powerful technique for forensic science. The electron microscope can achieve a wide range of magnifications simply by causing the electron beam to impinge on either a very small area (high magnification) or relatively large area (low magnification) of the specimen. Therefore X-rays can be generated from either a very small area ($<1\,\mu m$ diameter) or a relatively large area, the choice being controlled readily by the analyst depending on the analytical task at hand. In analytical range SEM–EDX is superior to XRF. The electron beam does not penetrate very far into the specimen, it is therefore possible to examine thin objects without receiving a signal from the mounting substrate, or examine individual layers of laminated materials such as paints without receiving a spectral contribution from lower layers. Therefore, with respect to spatial resolution of the specimen, SEM–EDX is superior to XRF.

A criticism often unfairly leveled at SEM–EDX is that the technique has a high limit of detection. If the specimen is a homogeneous mixture, then the limit of detection for any one particular element is of the order of a few tens of parts per million. Many atomic spectroscopic techniques, such as those based on ICP or AAS, or even XRF, have much better limits of detection than this. However, SEM–EDX still finds widespread application in forensic science simply because specimens are usually not homogeneous. Consider a tape lift a few square centimeters in size containing ten firearm cartridge primer particles each about $10\,\mu m$ in size. Each would weigh approximately a few nanograms, so the specimen would contain about 50 ng of analyte, say lead, barium and antimony, in total. If SEM–EDX was used at low magnification to analyze the entire tape lift it is very unlikely that any of the analyte would be detected, the particles are just too widely distributed over the tape. However, the actual particles contain very high percentage concentrations of the heavy metals, each element well above the detection limit for SEM–EDX. If the particles can be found using the microscope, and the

electron beam made to dwell only on the individual particles by adjustment of the magnification, then each element present will be readily detected. Fortunately a phenomenon associated with electron microscopy makes it easy under certain circumstances to locate particles such as those from cartridge primers. Not only does electron impact cause the specimen to emit X-rays, it also results in an output of electrons with various characteristics. Some of these secondary electrons are a result of inelastic collisions between the electron beam and atomic electrons in the specimen; the electron microscope captures secondary electrons from the specimen in order to produce the standard magnified image that we associate with SEM. In addition, some electrons from the beam collide with nuclei of atoms in the specimen and are bounced straight back out of the specimen with little loss of energy. These backscattered electrons can also be detected and used to create a magnified image of the specimen. Heavy nuclei within the specimen are more likely to produce backscattered electrons, therefore bright areas in the backscattered image correspond to areas rich in heavy elements. With appropriate signal processing the backscattered electron image can be tuned to show only elements of high atomic weight. In the case of the tape lift described above, the backscattered image can be tuned to reveal heavy elements such as lead, while light elements remain invisible.

Under these circumstances searching for primer particles is simple, especially if automation is used. Once particles are found by their bright backscatter signal, they can be analyzed readily. In principle, SEM–EDX could be used to analyze a minute particle in an enormous matrix so long as there is enough contrast between the particle and the matrix to enable it to be found. Under these circumstances SEM–EDX is capable of extremely low limits of detection.

In order to analyze the tape-lift using atomic spectroscopy, it must be digested in solvent, say 1 ml of acid. Using this approach the technique must be able to detect $50\,\mathrm{ng\,ml^{-1}}$ (50 parts per trillion) for it to be successful. It is unlikely that such limits of detection could be practically achieved for routine work. Furthermore, if results could be obtained using atomic spectroscopy, then interpretation would be difficult. Let us assume that the results indicate low levels of lead, barium and antimony. Although this could mean that the specimen contains mixed lead, barium and antimony particles characteristic of firearms primer, it could also indicate the presence of a mixture of pure lead particles, pure antimony particles, and pure barium particles, a situation which does not point to the presence of gunshot residue particles.

It is possible to analyze the tape lift for primer particles using XRF, but it is tedious and not efficient. First the technique does not have an equivalent to a backscatter image, therefore tiny primer particles are very hard to see, and the specimen must be searched by eye manually. Second, if a particle is found it will likely occupy only a small percentage of the volume of interaction of the X-ray beam, therefore analytical results will reflect the composition of the surroundings more than the particle itself.

The tape lift described above is an example of a specimen exhibiting extreme heterogeneity. Other important evidentiary materials are likely to be more homogeneous, but nevertheless SEM–EDX can still be valuable. A good example is paint comparison. Under low magnification a sample of paint appears to be quite homogeneous. SEM–EDX at this magnification gives results analogous to those obtained by techniques such as atomic spectroscopy or XRF, that is, a breakdown of all the elements contained within the specimen down to the limit of detection. SEM–EDX can yield further information that cannot be obtained using the other techniques, however. Consider the sample of paint discussed above in relation to XRF; the presence of Ca, S and Ba are detected. Analysis on the gross scale does not allow differentiation between paints filled with calcium sulfate and barium sulfate, or barium sulfate and calcium carbonate, or a mixture of those three compounds. However, with SEM all that is required is to zoom the magnification until individual particles are visible, and then analyze those particles individually. If particle analysis indicates Ca and S to be present in the same particle, then it can be safely concluded that calcium sulfate is present, particles containing only Ca indicate calcium carbonate, particles containing Ba and S indicate barium sulfate. In order to facilitate this process further the backscatter signal can be tuned to highlight particles containing Ba.

Similar to the case described for XRF, SEM–EDX suffers from a lack of specificity with respect to oxidation state of complex anions. For example the technique cannot distinguish between chloride, chlorate, perchlorate or hypochlorite, nor between sulfur, sulfide, sulfate, or thiosulfate.

Although SEM–EDX has a few shortcomings, it is a very versatile and efficient technique for elemental analysis. Although it suffers from an apparently high limit of detection, the ability to analyze submicrometre particles or specimen areas compensates to a great extent. If a particle or unusual region in the specimen can be found using the imaging options, then it can be analyzed, giving the technique an extremely low working limit of detection. Furthermore, the instrument allows imaging of specimens at high or low magnification with very high depth of focus. This facility can provide useful morphological

information relating to trace evidence, such as fiber cross-section and surface morphology. Finally, an important consideration is that often the specimen needs little or no treatment before analysis, and the technique is not destructive.

Mass spectrometry

A characteristic of substances is the atomic weight or molecular weight of its constituent atoms or molecules. Atoms or molecules are too small to be weighed directly by conventional equipment such as a microbalance, but their weight can be measured indirectly using a device called a mass spectrometer. In forensic science, mass spectrometry is mainly used for molecular analysis, therefore this section will deal only with this topic.

Before mass measurement, molecules within the specimen are raised to a charged state of high potential energy. The classical approach to this is to bombard the specimen with a beam of electrons; this technique is referred to as electron impact ionization. The excitation might produce a state of such high potential energy that molecules within the specimen break up into fragments. This outcome sounds counterproductive, but in reality it can yield useful information. Fortunately, fragmentation is not a random process, it occurs at the weakest bonds within compounds, or at locations where stable fragments are produced.

Figure 6 shows the mass spectrum produced by a sample of cocaine.

The peaks in the spectrum indicate fragments detected; the horizontal axis indicates the weight of the fragment detected, and the vertical axis indicates the abundance of those fragments detected. Weight displayed is a relative measure. Mass spectrometers are calibrated using the assumption that carbon atoms have a weight of 12. On this scale hydrogen atoms have a weight of 1, oxygen 16, nitrogen 14, sulfur 32 etc. In reference to Fig. 6, the peak at 303 is due to intact cocaine molecules that have not fragmented and therefore contain 17 carbon atoms, 21 hydrogen atoms, four oxygen atoms and one nitrogen atom

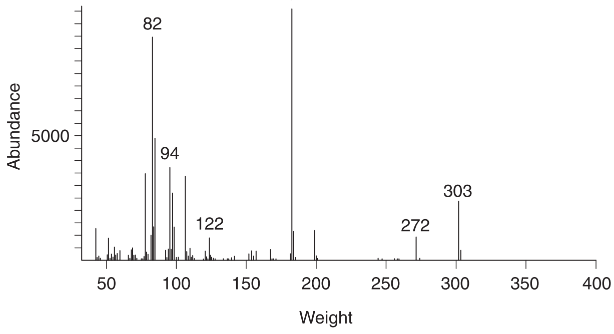


Figure 6 Mass spectrum for cocaine.

$(303 = 12\times17+1\times21+16\times4+14)$; this peak is referred to as the molecular ion. Other peaks relate to fragments derived from the molecule, for example the peak at 182 is due to a fragment containing 10 carbon atoms, 16 hydrogen atoms, two oxygen atoms and one nitrogen atom $(182 = 12\times10+1\times16+16\times2+14)$.

The fragmentation pattern produced by a given molecule is distinctive, and in some instances unique, it is therefore a good means of identifying compounds present in a substance. However, some collections of atoms very strongly direct fragmentation, with the result that various molecules that contain those groups produce fragmentation patterns that resemble each other. For example, the $—CH_2—NH_2$ group in primary amines very strongly directs fragmentation, with the result that, as a group, the amphetamines all produce mass spectra with strong peaks at 44, and few other fragments. With respect to compounds of forensic interest, mass spectrometry can fail to produce good discrimination, within a class, for substances such as aliphatic or aromatic hydrocarbons (as found in accelerants), nitrate esters (explosives), as well as amphetamine analogues and methylamphetamine analogues.

The cause of this lack of specificity is that electron impact ionization raises molecules to a state of very high potential energy. It is possible to use special excitation techniques that do not encourage fragmentation, such as chemical ionization, fast atom bombardment, laser desorption or electrospray ionization. Instead of peaks arising from fragments of molecules, mass spectra acquired using these techniques give a peak corresponding to the molecular weight of the compound. Using the example of amphetamine, the chemical ionization spectrum of this substance shows a strong peak at 135 instead of a strong peak at 44 as obtained using electron impact ionization. Therefore, although the chemical ionization spectrum is still very simple, in this example it is more discriminating than the electron impact spectrum.

Mass spectrometry is, therefore, a very versatile technique. The fragmentation pattern generated by electron impact in most instances allows good discrimination between compounds. In those instances when fragmentation occurs too readily in one direction, other ionization techniques can be brought into action.

The main limitation of mass spectrometry is that if the specimen is a mixture of compounds, then the mass spectrum acquired will be also be a mixture. Unless the task is target analysis, this can make interpretation of data and detection of trace level components very difficult. This shortcoming can be overcome by using a technique known as mass spec-

trometry–mass spectrometry (a discussion of which is beyond the scope of this chapter), or by combining mass spectrometry with a chromatographic technique such as gas chromatography, liquid chromatography or capillary electrophoresis.

See also: **Analytical Techniques**: Spectroscopy: Basic Principles; Mass Spectrometry; Hyphenated Chromatographic-Spectroscopic Techniques.

Further Reading

Chapman JR (1985) *Practical Organic Mass Spectrometry*. London: Wiley.

Colthup NB, Daly LH and Wiberley SB (1990) *Introduction to Infrared and Raman Spectroscopy*. San Diego: Academic Press.

Davis R and Frearson M (1987) *Mass Spectrometry*. London: Wiley.

Dzunikowski B (1989) Energy dispersive X-ray fluorescence analysis. In: Srehla G (ed.). *Comprehensive Analytical Chemistry*, Vol. XXIV. Netherlands: Elsevier.

Goldstein JI, Newbury DE, Echlin P, Joy DC and Fiori C (1981) *Scanning Electron Microscopy and X-ray Microanalysis*. New York: Plenum.

Goodhew PJ and Humphreys FJ (1988) *Electron Microscopy and Analysis*, 2nd edn. London: Taylor and Francis.

Hesse M, Meier H and Zeeh B (1997) *Foundations of Organic Chemistry: Spectroscopic Methods in Organic Chemistry*. New York: G Thieme Verlag.

Humecki HJ (1995) *Practical Spectroscopy*, vol. 19, Practical Guide to Infrared Microspectroscopy. New York: Marcel Dekker.

Lajunen LHJ (1992) *Spectrochemical Analysis by Atomic Absorption and Emission*. Cambridge: Royal Society of Chemistry.

McLafferty FW and Turecek F (1993) *Interpretation of Mass Spectra*, 4th edn. Mill Valley, USA: University Science Books.

Moffat AC (ed) (1986) *Clarke's Isolation and Identification of Drugs in Pharmaceuticals, Body Fluids, and Post-mortem Material*. London: Pharmaceutical Press.

Reed SJB (1975) *Cambridge Monographs on Physics: Electron Microprobe Analysis*. Cambridge: Cambridge University Press.

Robertson JR (ed) (1992) *Forensic Examination of Fibres*. London: Ellis Horwood.

Roush PB (eds) (1987) *The Design, Sample Handling, and Applications of Infrared Microscopes*. Philadelphia: American Society for Testing and Materials.

Skoog DA and Leary JJ (1992) *Principles of Instrumental Analysis*, 4th edn. Philadelphia: Saunders College Publishing.

Watson JT (1985) *Introduction to Mass Spectrometry*, 2nd edn. New York: Raven Press.

Willard HH, Merritt LL, Dean JA and Settle FA (1981) *Instrumental Methods of Analysis*, 6th edn. New York: Van Nostrand.

Williams DH and Fleming I (1995) *Spectroscopic Methods in Organic Chemistry*, 5th edn. London: McGraw Hill.

Williams KL (1987) *An Introduction to X-ray Spectrometry*. London: Allen and Unwin.

Yinon J (ed) (1987) *Forensic Mass Spectrometry*. Boca Raton: CRC Press.

Spectroscopy: Basic Principles

K P Kirkbride, Forensic Science Centre, Adelaide, South Australia

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0411

Introduction

Spectroscopy is an analytical technique that uses electromagnetic radiation to probe the atomic or molecular composition of substances. This article outlines the basic principles of spectroscopy for non-specialists.

Light and Electromagnetic Radiation

Light is a phenomenon that is a special case of a much wider phenomenon called electromagnetic radiation. As the name indicates, electromagnetic radiation is a travelling disturbance in space that comprises electric and magnetic components. Unlike the field associated with a common magnet or the earth, the magnetic field associated with electromagnetic radiation is constantly changing its direction and strength. The electric field component undergoes exactly the same behavior. **Figure 1** illustrates what would be observed if it was possible to see the electric and magnetic

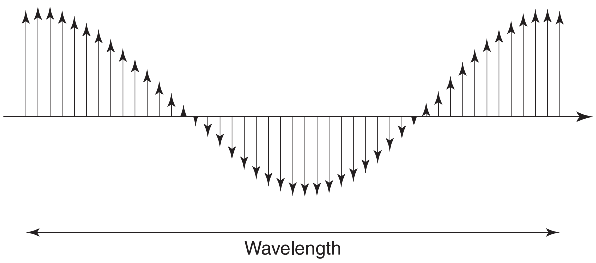


Figure 1 Representation of a ray of electromagnetic radiation traveling through space, left to right. The small vertical arrows represent the strength and direction of the electric field component of the radiation at discrete locations (the magnetic field behaves in an analogous manner). The disturbance is periodic, with the distance between field components of identical strength and direction being the wavelength of the radiation.

components of a ray of radiation as it passes. It is the interaction of the electric and magnetic components of radiation with matter that are responsible for all the phenomena we associate with radiation, such as human vision, sunburn, microwave cookery, radio and television, and of course spectroscopy.

There are some important fundamental features of electromagnetic radiation. The periodicity of the disturbance, or the distance between adjacent troughs or adjacent peaks, is referred to as the wavelength of the radiation. The number of times the electromagnetic field undergoes a complete cycle per unit time is called the frequency. The speed with which radiation can travel is limited, and moreover, radiation of all wavelengths travels at a single speed. This is referred to as the speed of light, universally denoted by the symbol *C*. The speed of light (and every other radiation) is dependent on the substance through which it travels, and reaches a maximum in a vacuum. The speed of light is a constant of proportionality between frequency and wavelength as described by equation 1.

$$C = \nu \times \lambda$$

where ν is radiation frequency, λ is radiation wavelength and C is speed of light.

The interpretation of this equation is that wavelength and frequency are related. Electromagnetic radiation of long wavelength must have a low frequency, whereas radiation of short wavelength must have a high frequency. At the long wavelength (low frequency) end of the electromagnetic radiation spectrum are radio waves, at the other is high frequency (short wavelength) radiation such as gamma rays. Between these extremes lie other forms of radiation, including microwaves, infrared, ultraviolet, and that narrow band of radiation that can be directly observed by humans, visible light.

Radiation carries energy with it, a phenomenon which is made use of in the domestic microwave oven, for example. The amount of energy carried is proportional to the frequency of the radiation. As frequency and wavelength have an inversely proportional relationship, energy carried is inversely proportional to wavelength. Radiation of high frequency and short wavelength (e.g. x-rays, gamma rays) therefore is of high energy, whereas low frequency radiation which must have long wavelength (e.g. radio waves) carries little energy.

Matter

Given that radiation is an electromagnetic phenomenon, the main feature of matter as far as its interaction with radiation is concerned is the distribution of its electric and magnetic components.

From the subatomic scale to the molecular, matter exhibits a variety of electric and magnetic features with which radiation can interact. **Figure 2** depicts some of these features for a molecule of ammonia. The forces that hold the molecule together arise from attraction between negatively charged electrons shared between the positively charged nuclei of the nitrogen and hydrogen atoms. The nuclei of each of the hydrogen atoms contain a single charged particle, a proton. The share of electrons between the hydrogen nuclei and the nitrogen nucleus is not equal; nitrogen commands a larger share. Each nitrogen–hydrogen bond therefore has a heterogeneous distribution of charges resulting overall in a positively charged end (towards the hydrogen atom) and a negatively charged end. Such a bond is described as having a dipole moment or being polarized. As each nitrogen–hydrogen bond has a similar polarization there is an additive effect with the result that the entire molecule is polarized or has a dipole moment.

Each nucleus of the hydrogen atoms is comprised of a single positively charged proton, each acts like a small bar magnet, or is said to possess a magnetic moment. Potentially the entire ammonia molecule, its bonds, its protons, and its electrons each have the necessary prerequisites to interact with radiation.

Interaction Between Radiation and Matter

In the previous section it was shown that matter can be thought of as possessing many minute electric and magnetic moments on the microscopic sale. Radiation, which is also electromagnetic in nature, therefore potentially can interact with matter on the microscopic scale. If interaction is to be significant, however, there has to be a match between the frequency of the electromagnetic oscillations executed

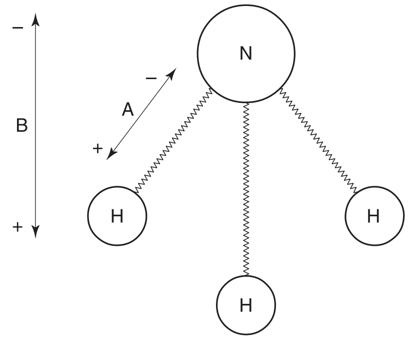


Figure 2 Electrical moments in the ammonia molecule. 'A' represents the electrical field resulting from unequal sharing of electrons between nitrogen and hydrogen nuclei. 'B' represents the overall field resulting from combination of three 'A' moments.

by the radiation and oscillations of the electromagnetic moments within the matter. The situation is analogous to that of pushing a person on a swing. Say the person on the swing completes a cycle once every second. Efficient transfer of energy takes place when pushes are in unison with the motion of the swing (i.e. with a frequency of one per second). If the frequency of pushing is three per second, then two of the three will accomplish nothing, energy transferral is not efficient.

Spectroscopy is a tool that makes use of interactions between matter and radiation. The outcome of the interaction allows scientists to infer analytical information regarding the atomic or molecular composition of the matter.

Spectrometers

The interaction between radiation and matter can lead to the occurrence of a variety of phenomena; two important ones as far as spectroscopy is concerned are absorption and emission.

An appreciation of absorption can be gained by considering the properties of the lens in a pair of sunglasses. Through the lens the image of the outside world is duller, some of the radiation (light) encountering the lens does not pass through – the lens absorbs it. The basis of absorption spectroscopy is the measurement of the manner in which matter absorbs radiation.

If a substance absorbs radiation it can re-emit it in a modified form; this is referred to as emission. The basis of emission spectroscopy is the measurement of the manner in which matter emits radiation. The output from measurement of either emission or absorption is called the spectrum of the substance.

Figure 3 depicts the important features of simple instrumentation that can be used for absorption spectroscopy, and a typical spectrum. Although all absorption spectrometers might not be exactly described by Fig. 3, as far as this article is concerned the important issue is the process leading to the spectroscopic output, and the features of the output itself, rather than the exact mechanism by which it was generated.

Three basic elements are present: a radiation source; a variable 'filter'; and a detector. It is usual for the source to emit radiation of many frequencies within a region of the entire electromagnetic spectrum, for example the infrared region, the visible light region, x-rays, etc. The region over which the source operates in a given spectroscopic technique is often used as a means to name that technique. For example, if an infrared source is used, then the technique is referred to as infrared spectroscopy.

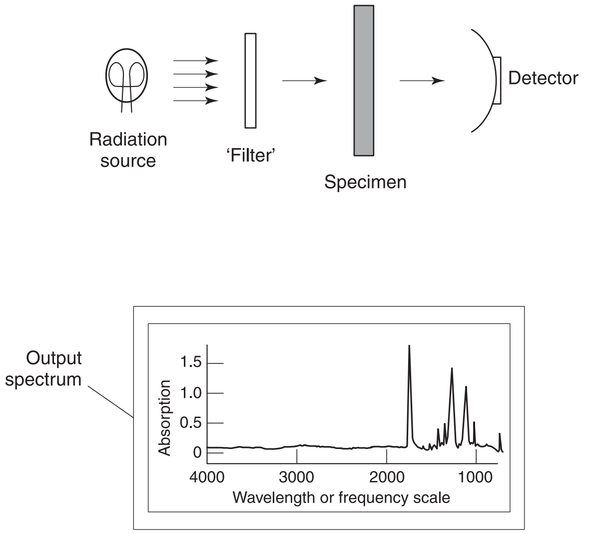


Figure 3 Schematic diagram of a spectrometer. Radiation from the source is passed through a 'filter', then the specimen, and then on to the detector. The spectral output is a plot showing the range of frequencies (or wavelength) that strike the specimen, versus the extent of absorption of each frequency.

The filter selectively modifies radiation from the source, ideally it should allow radiation of only one frequency at a time to pass through. The radiation from the filter is then allowed to pass through the specimen to be analyzed. The detector then collects the radiation that is transmitted by the specimen, where its intensity is measured. Spectroscopic measurement of the specimen is performed by 'sweeping' the filter through its entire frequency range allowing (ideally) the detector to measure the radiation transmitted by the specimen one frequency at a time. A graphical representation of measurements made by the detector is the spectrum of the specimen; typical output is shown in Fig. 3.

The horizontal axis in the spectrum defines the range of frequencies of the radiation that strikes the specimen. The vertical axis indicates how much of each frequency is absorbed by the specimen (i.e. a peak at a certain frequency indicates that the specimen absorbs at that frequency).

The ideal of spectroscopy is that because different substances undergo different motions on the microscopic scale, then the spectrum of each must have a unique pattern of peaks. The reality is that substances with very similar structures could behave in a very similar manner, with the result that instrumentation is not capable of discriminating them.

Throughout this article the term spectrometer has been used to describe the device that acquires a spectrum, whereas spectroscopy has been used to describe the technique. This usage is common in

modern publications, but it is by no means the only terminology that might be encountered. The words spectrograph, spectrophotometer, and spectroscope are used to describe apparatus, whereas spectrometry and spectrophotometry are used interchangeably with spectroscopy.

See also: **Analytical Techniques**: Mass Spectrometry; Spectroscopic Techniques; Hyphenated Chromatographic-Spectroscopic Techniques.

Further Reading

Banwell CN (1972) *Fundamentals of Molecular Spectroscopy*, 2nd edn. London: McGraw-Hill.

Barrow GM (1962) *Introduction to Molecular Spectroscopy*. New York: McGraw-Hill.

Brown JM (1998) *Molecular Spectroscopy*. Oxford: Oxford University Press.

Denney RC (1982) *A Dictionary of Spectroscopy*, 2nd edn. London: Macmillan.

Hesse M, Meier H and Zeeh B (1997) *Foundations of Organic Chemistry: Spectroscopic Methods in Organic Chemistry*. New York: G Thieme Verlag.

Pavia DL, Lampman GM and Kriz GS (1979) *Introduction to Spectroscopy: a Guide for Students of Organic Chemistry*. Philadelphia: WB Saunders.

Straughan BP and Walker S (1976) Spectroscopy. London: Chapman Hall.

Williams DH and Fleming I (1995) *Spectroscopic Methods in Organic Chemistry*, 5th edn. London: McGraw Hill.

Animals *see* **Anthropology**: Animal Effects on Human Remains.

ANTHROPOLOGY

Contents

Overview
Animal Effects on Human Remains
Archaeology
Assessment of Occupational Stress
Bone Pathology and Antemortem Trauma in Forensic Cases
Determination of Racial Affinity
Excavation and Retrieval of Forensic Remains
Morphological Age Estimation
Sex Determination
Skeletal Analysis
Skeletal Trauma
Stature Estimation from the Skeleton

Overview

W M Bass, University of Tennessee, Knoxville, TN, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0415

Introduction

Forensic anthropology is that field of science whose major focus is the identification of more or less skeletonized remains, either human or animal, in a legal context. Forensic anthropology is a rapidly growing subarea of anthropology that is an out-

Estimation of stature from the long bones A number of formulas exist that enable calculation of stature from long bones (bones from the arms and legs). In general most calculations are based on the maximum length of the bone. Formulas exist for most racial and ethnic groups.

Length of time since death There are many factors involved in determining the rate of decay, but the major factor is temperature. Decay is much faster in summer (warm conditions) than in winter.

Court Testimony

Occasionally, the forensic anthropologist will be called upon to testify in a court case concerning his/her professional opinions about various aspects of the investigation. The forensic anthropologist usually testifies as an expert witness. Most forensic anthropologists testify for the prosecution (since they work for state or federal agencies), although in recent years, as the knowledge of expertise in forensic anthropology has spread in the scientific and legal communities, their advice is sought by both sides in the medicolegal field.

International development

The publication in 1978 of *Forensic Fetal Osteology* by Fazekals and Kosa was the beginning of considerable research on fetal bones, not only in Hungary, but in all Europe and the rest of the world. Research in forensic anthropology has greatly increased since the 1970s. An example is the publication in 1987 of *Death, Decay and Reconstruction* edited by Boddington, Garland and Janaway. Janaway has been a leader in England into the relationship between cadaver decay and death scene materials. An example of international research is the publication of Capasso, Kennedy and Wilczak in 1999 of an *Atlas of Occupational Markers* where researchers in the USA and Italy have combined and published their research interest.

Many articles in forensic anthropology have been published by researchers in India and Japan. Research in Australia by Walter Wood MD similar to that by Janaway in England and Bass in America has been reported on by Donna MacGregor who won the prize for the best student paper presented at the 1999 American Academy of Forensic Sciences meeting in Orlando, FL.

The American Academy of Forensic Sciences has assumed a leadership role in encouraging international research and publication in its *Journal of Forensic Sciences*.

Conclusion

When human bones are found the professional called upon to identify those skeletal remains is the forensic anthropologist. Forensic anthropology is the most recent subarea to develop from physical anthropology. The forensic anthropologist applies the skills and techniques learned in the study of prehistoric populations to modern forensic cases. He applies the knowledge to establish the age, sex, race or ancestry, stature, handedness, and often the manner of death of modern skeletal remains found in a medicolegal context. He is occasionally called upon to testify in court as an expert witness. The growth of forensic anthropology has been worldwide with major research and publication occurring in Australia, England, Hungary, India, Italy, Japan, The USA and to a lesser extent in other countries of the world.

See also: **Anthropology**: Archaeology. Skeletal Analysis, Morphological Age Estimation, Sex Determination; **Education**: An International Perspective. **Expert Witness**: Qualifications and Testimony. Determination of Racial Affinity.

Animal Effects on Human Remains

M Y İşcan and B Q McCabe, Adlı Tip Enstitusu, Istanbul Ünivesitesi, Istanbul, Turkey

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0762

Introduction

Assessment of trauma is one of the most difficult tasks of forensic pathologists. Cause may be related to accidents, homicide or suicide, or may be natural. Whatever the manner, the source for the trauma must be sought. In general there are some clear signs, such as a stab wound, fall from a height, ligature for hanging or a bullet hole. There are also less clear signs, such as a fatality caused by an animal. Animal attacks can be, and have been, misinterpreted as heinous acts.

Perhaps the most outstanding example of how an animal can cause a fatality to humans comes from Africa. According to an anthropologist, a leopard had punctured an australopithecine skull. Taphonomic explanation of the event is that the human fell prey to the cat and was carried up into a tree, where the leopard habitually consumed its prey at its leisure. In

order to carry the carcass, the animal had to penetrate the victim's skull with its canines. The fossil skull, found millions of years later, shows these punctures, the dimensions of which are similar to the distance between the right and left canines of both jaws of the predator.

Wild animals can cause injuries, especially carnivores, such as lions, tigers, wolves, coyotes, bears, cougars, alligators and crocodiles, living near human settlements. Docile animals, such as horses and cattle, can also deliver severe or fatal wounds. Domestic animals, such as chickens, ducks and geese, have also been known to inflict serious wounds. Still others, such as marine animals, can render severe damage or kill through predation or scavenging.

Lesions caused by animals must be assessed in order to resolve any question as to whether or not all trauma is from sources like that of a knife. This is particularly important if stab wounds are next to a place bitten off by an animal. Sometimes the question arises as to whether or not the injury occurred ante-, peri- or post mortem. Since carnivores scavenge as well as hunt prey, injuries to bodies may be the result of an opportunistic meal due to a homicide, accident or suicide. Carnivores are not the only animals to scavenge. Rodents, insects, crabs, fish and fowl are not likely to pass up a free meal.

Dogs and Cats

The animal that most often causes trauma to humans is the domestic dog. The family pet is responsible for approximately 88% of all bite wounds. This figure is probably a conservative estimate because many bites go unreported. Fatalities due to dog bites are rare. A total of 266 deaths were recorded in the United States between 1975 and 1994. Attacks occur most often among those least able to defend themselves, i.e., children less than 10 years of age and the elderly.

Over 70% of injuries are on the head, neck and face. Lower limbs and upper extremities are the next most common sites. **Figure 1** shows the location of common dog bite injuries. The size and age of the victim determine bite sites. Adults tend to be bitten on the extremities, children on the face, head and neck, and children under 2 years of age are particularly susceptible to cranial injuries. The injuries caused by dogs are of three basic types: crush wounds, puncture wounds, and avulsion wounds. Depending on the breed and size of dog, the animal is able to exert a biting force of up to 12 420 kPa (1800 $\mathrm{lbf\,in^{-2}}$). A crush wound is the result of this biting force. Death by asphyxiation has been noted in some cases, caused by crushing bites to the neck. Puncture wounds are caused by the canine teeth. These wounds are seen on the skulls of small children and infants. They are also found in other areas of the body such as the extremities. Multiple puncture wounds are not uncommon. Avulsion wounds are those that remove flesh. Typically these are made when the animal shakes its head during the attack, thereby tearing the skin and muscle tissue. Attacks generally involve a combination of biting, clawing and crushing forces, resulting in wounds with a characteristic pattern of punctures, lacerations and avulsions of skin and soft tissues. Angry dogs move their heads vigorously from side to side as they bite, tearing soft tissues and sinking the teeth deeper into the victim.

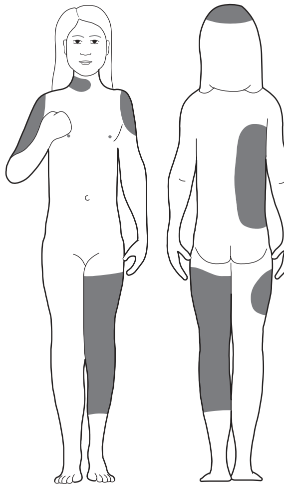


Figure 1 Common distribution of dog bite injuries. (From Lauridson and Myers, 1993.)

Although not as common, dogs sometimes inflict nonbiting wounds. The force of an attack can cause impact injury resulting in broken bones. In one example a kennel owner, while bending to unlock a kennel door, was hit in the throat by a greyhound. The cause of death was determined to be blunt force trauma to the throat.

Another aspect that should be looked at is attack by multiple animals, or 'pack attack'. Dogs are social animals and possess an instinct toward pack association. Wild and/or feral canines tend to live and act in territorial groups. Territorial animals display strong defensive behavior and social dominance in the home areas. Assaults by several animals have a greater probability of resulting in fatal injuries than attacks by a single animal. Many more injuries are inflicted and the attack usually ends in the death of the prey.

While domestic cats may attack humans, rarely are wounds more severe than a small puncture wound or a scratch. However, in the presence of a deceased person a common house cat can and may opportunistically attack a body. Damage is determined by the time the animal has been in the presence of the body. Usually, the soft tissue is stripped from the exposed

areas, the most frequent sites being the face, head, neck and exposed extremities. The speed with which domestic animals are able to damage a corpse is surprising. In just a few days, and in some cases a matter of a few hours, injuries can be delivered that can, and sometimes do, confuse the determination of the cause of death, thereby leading investigators to suspect a brutal and vicious homicide has occurred.

Rodents

Free-range indoor domestic pets, such as gerbils, mice, rabbits, hamsters, guinea pigs, etc., can cause postmortem damage to human bodies when they come into contact with them. Postmortem scavenging is common among rodents and other domestic pets. This damage can pose many problems for the death investigator. Several species of wild rodents are well known for tendencies to collect bone, scattering and destroying body parts and crime scenes. When allowed to roam freely throughout a dwelling, domestic rodents will also collect items such as bones, skin, nails etc. and return them to their nests. The damage caused by these animals can significantly alter or destroy indicators of cause of death and of identification.

Rodents possess large incisor teeth that continuously grow throughout the animal's life. Pet stores sell products to place in the animals' cages to help them wear down the teeth and to keep them sharp. If the animal is not able to gnaw and wear down the incisors, the teeth will continue to grow and the animal will die from the inability to feed itself. It has been suggested that these animals scavenge human remains for the purpose of keeping the teeth at a usable length. Most rodents leave characteristic gnaw marks (**Fig. 2**). These marks are often, though not exclusively, seen as parallel stria or channels, depending on the chewing behavior. The areas prone to gnawing are typically the exposed or unclothed parts of the body. Damage to the soft tissues is seen in a layered effect from the skin down to bare bone. The margins of the injury are marked by a series of crenations. Small, slit-like wounds on the soft tissues, caused by the animal's bite, are observed in pairs.

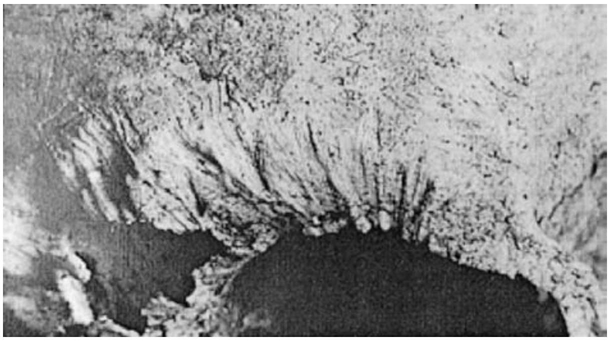


Figure 2 Left supraorbital margin showing classic rodent gnawing. (From Haglund, 1992.)

Rodent scavenging has been mistaken for a scalping injury in one case. A female was found dead in her apartment, lying on the floor in a supine position. Her face showed an extensive soft-tissue wound of the forehead, nose and upper lip. During autopsy, finely crenated edges were discovered at the wound sites. Multiple small, paired cuts were observed in the area of the wings of the nose and the upper lip. These wounds led investigators to search the victim's home for a rodent. An open pet cage was located and subsequently the occupant, a golden hamster, was also found. Several small pieces of the deceased's skin; fatty tissue and muscle were recovered from a nest made of paper towels in an open drawer.

Chickens and Birds

A variety of birds are kept by people as pets or as income- or food-producing stock. Injuries caused by these creatures are rarely seen in living persons, and even more rarely as serious injuries resulting in death. Nowadays the injuries stem mostly from domestic animals, such as parrots, parakeets, finches, etc., which nip a finger of an uncautious person. Some injuries must occur in and around the raising of these animals but there are no records.

The forensic literature mostly records postmortem lesions caused by crows, owls, buzzards or seagulls. These injuries exhibit little stab–like or characteristic puncture wounds resulting from the hard bills. In 1898 Vonn Hoffman gave a description in the *Altas of Forensic Medicine* of a child's corpse found under a yew tree in the garden of Belvedere Castle in Vienna. His description of the injuries included both the anterior and posterior side of the lower part of the body. The wounds show multiple, yet parallel, sharp marginal slits, 3–4 mm long. The wounds mainly divided the skin but here and there involved the subcutis and reached down as far as the muscles. The species of bird(s) that had caused the lesions was not mentioned.

Attacks by buzzards and seagulls have been reported and/or observed. The birds generally attack by diving at the victim from behind. Attacks have been known to continue for as long as the person was in motion. Injuries to the scalp are caused by the talons and beaks of the animals and have reported to be up to 14 cm in length.

Chickens are capable of inflicting deep wounds with their beaks and their claws. The claws are able to puncture deep into the victim's flesh, resulting in painful lesions. They can also use their beaks as effective weapons. The wounds caused by the bills

have a characteristic hourglass shape. The pushing and pulling action of pecking tears the skin as well as puncturing it.

Vultures and buzzards are excellent scavengers. Animal remains are often completely defleshed in a matter of hours by these efficient foragers. Damage to bone occurs from the stripping and tearing action of the beaks and talons. Some birds, for example the turkey vulture, have beaks that are weak and are unable to dine on a carcass until decomposition has begun and the tissues are softened.

Crows and hawks play a role in scattering bones. Their hard beaks are capable of leaving small punctures and scratches on bone surfaces.

Farm Animals

People live and work around large animals, such as cows and horses, daily. Traumas must occur regularly, however there is little information describing the types of injuries. Considering the size difference between humans and these mammals one would imagine that the injuries that occur must be of the blunt force, crushing variety. Some of these injuries are listed in **Table 1**. In one reported case, a farmer was found dead in his stable. An autopsy revealed a depressed fracture of the skull (**Fig. 3**). The mandible was shattered into large fragments and a gaping, exploding fracture line ran throughout the right petrous bone and the pituitary fossa and into the anterior region of the left middle cranial fossa. The farmer had been pushed against the wall by a cow; and the cow's left horn hit the right parieto-occipital region of the man's head, crushing it against the stable wall. The concurrent pressure from the head of the cow caused the lower jaw to break into pieces.

Injuries from horses and horseback riding have been reported throughout history. The first pattern of injury discussed here occurs while riding a horse and involves either a fall from the horse or the horse falling and crushing the rider. For example, the horse

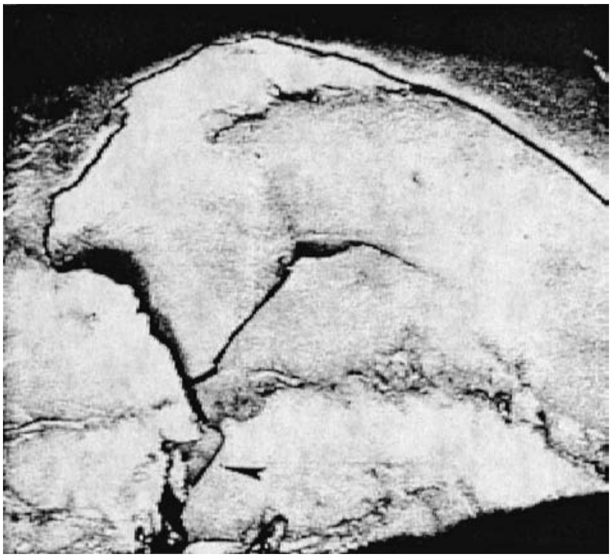


Figure 3 Depressed skull fracture from a cow horn. (From Rabl and Auer, 1992.)

rears back and falls over, landing on the rider and causing a severe crush injury. The rider may also be thrown from the back of the animal, resulting in a skull fracture or severe spinal damage. Paralyses resulting from head and neck injuries are major concerns. The second form may involve people who were on the ground in close proximity. The horse may kick or crush an individual while it is being groomed, shod or fed. Broken legs are not uncommon. These injuries are classified as bluntforce mechanical trauma. Rodeo participants, whether riding or acting as one of the many behind-the-scenes animal handlers, often sustain injuries inflicted by bucking animals, or crushing injuries as described in the case above.

Wild animals

People come into contact with wild animals in a number of ways. Interest in personal fitness and recreation sends more and more people outdoors, and population increases deplete the habitats of indigenous

Table 1 Cases of fatal injuries caused by domestic animals from blunt force trauma

Person	*Age (years)*	*Event*	*Cause of death*
Man	60	Hit by a bull	Thoracic injuries
Man	58	Hit by a bull	Rupture of the inferior vena cava
Woman	72	Kicked by a cow	Fat embolism
Woman	71	Kicked by a cow	Pulmonary embolism
Man	39	Kicked by a cow	Rupture of the liver
Fetus	—	Mother kicked by a cow	Placental rupture
Child	1.8	Hit by a cow	Reflexive death (carotid artery)
Man	77	Hit by a sheep	Pulmonary embolism

Adapted from Rabl and Auer (1992)

animals. These activities sometimes place humans directly in contact with predators such as large cats, bears, wolves and coyotes. The attacked person may sustain injuries to the head, arm, chest and back. Wounds caused by dogs and similar carnivores may be puncture, crush and avulsion, as illustrated in **Fig. 4**.

Wolves sometimes attack humans and are also known as scavengers. For several decades archaelogists have been interested in the effects of carnivores on bone. Paleoanthropologists are beginning to reassess their earlier suspicions of cannibalism in favor of faunal scavenging. Forensic anthropologists are concerned with these bone alterations and their implications. In order to reassess canid scavenging and bone modification, a group of captive timber wolves and road-killed deer were investigated. This study indicated ways in which scavenging can obscure the cause of death. The animals are first interested in open wounds or blood oozing, as in the area of the nose or mouth. They lick, chew and possibly enlarge the area. Scavenging activities usually last several days. Following the focus on any wounds, the next areas of the body to be consumed are meaty portions of the arms, and the chest cavity is opened and emptied of its contents. Attention is then paid to the ends of the bones. Wolves can effectively destroy a body in several days.

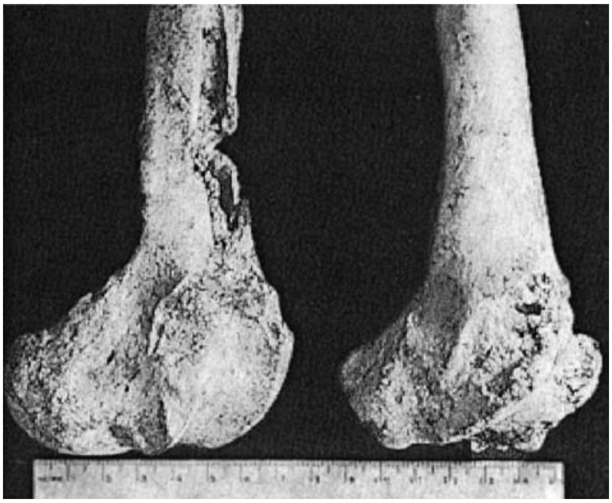


Figure 4 Distal femur showing punctures and furrowing of trochlea. (From Binford, 1981.)

Large cats, such as lions, tigers, cougars and leopards have been known to attack humans. These animals have strong jaws and teeth equipped with propioceptors or sensors that allow them to adjust their bite when the teeth encounter bone. They maneuver their teeth between their victim's vertebrae and into the spinal cord, and can actually realign their jaws so the teeth can enter the intervertebral space. Most of the attacks are directed at the neck and head of the prey. The food sources of these animals are usually prey with thick muscles at the nape of the neck, for instance wild cattle, or with horns or antlers protecting the nape. Consequently, they attack the side of the neck or throat. Jaguars and leopards have also been observed to bite the throat; lions and tigers may place a paw on the victim's back, then grasp the prey with their teeth, and drag the animal down. Tigers are also known to use a paw to suffocate their prey. A frequent slaying process for these animals is to shake the victim by the neck, resulting in hypertension injuries. Other injuries include multiple lacerations, abrasions, contusions and areas of avulsion, and punctures into the skull. **Table 2** provides a list of injuries caused by large cats. **Figure 5** shows extensive scalp lacerations and canine puncture.

Occasionally, encounters with bears prove fatal. Between 1973 and 1988 eigth deaths were attributed to bears in Alberta, Canada: four by grizzly bears (*Ursus arctos horribilus*) and four by black bears (*Ursus americanus*). Both of these animals are omnivorous. Of the two, the grizzly bear is the most dangerous. In the eight cases reported, there were six injuries to the head and neck, three to the face, six to the trunk, four to the upper limbs, and four to the lower limbs. Teeth, paws or a combination caused all injuries. Injuries in bear attacks are primarily directed at the head, then at the neck. The main weapons used by these animals were teeth. Contrary to popular opinion, claws are used only for rolling or moving a victim when the jaws prove insufficient.

Table 2 Examples of attacks by large cats

Person	*Age (years)*	*Event*	*Animal*	*Injury*
Male	28	Cleaning cage	Tiger	Neck, from animal attempt to suffocate
Male	18	Jogging	Mountain lion	Fatal injuries
Female	40	Crosscountry run	Mountain lion	Puncture wounds to skull, scratches from claws to upper extremities, bites to face
Female	30	Escaped animal	Jaguar	Fatal injuries
Female	2	Family pet	Leopard	Fatal injuries

Modified from Cohle *et al*. (1990), Rollins and Spencer (1995) and Wiens and Harrison (1996).

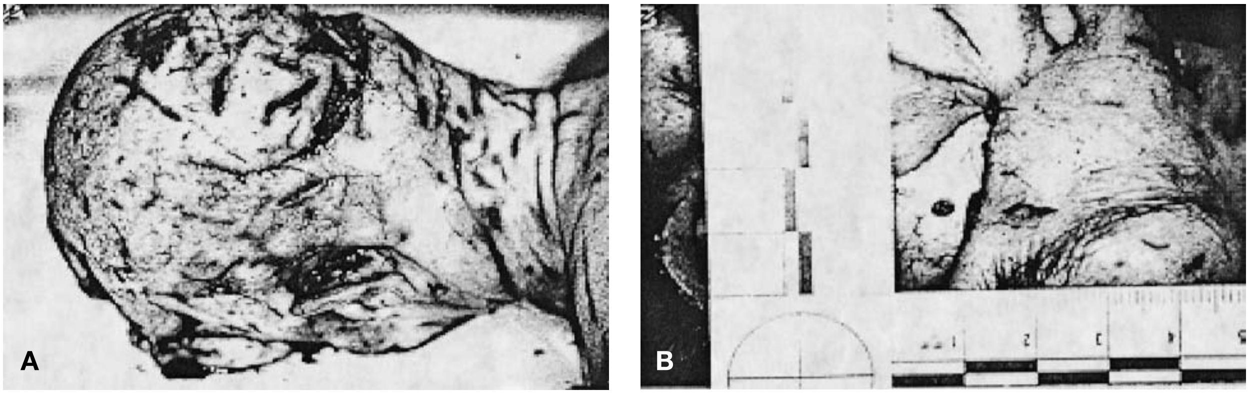


Figure 5 **(A)** Extensive scalp lacerations; **(B)** canine puncture into left occipital area. (From Rollins and Spencer, 1995.)

Reptiles

Other popular pets are nonvenomous giant constricting snakes of the Boidae family, such as the reticulated python (*Python reticulatus*). Authenticated reports of human deaths caused by constricting snakes are rare: a 1989 study attributed four deaths to these snakes in the United States since 1978. These animals kill by first securing their prey by biting and then wrapping its coils around the victim and constricting until the prey ceases moving. The dentition of (*P. reticulatus*) consists of two parallel rows of numerous recurved homodont teeth in the maxilla and one row in the mandible. Wounds caused by these teeth are small punctures arranged in a linear and parallel fashion. These bite marks were readily observed on the face and neck of one of the four confirmed victims. **Fig. 6** shows a snake skull with recurved teeth for securing prey. In three of the four cases the victim was sleeping when attacked by the snake. The fourth victim was asphyxiated while cleaning the animal's cage with the animal draped over the owner's shoulder.

Crocodiles and alligators cause considerable injury and fatality. These two animals are found in a large part of the world. Of the two reptiles, the crocodile is the larger and injuries it causes are more often fatal. Once they catch their prey, they rotate until part of the body of the victim, such as the head or appendages, is dismembered from the rest of the body. Some fatal cases include decapitation, loss of limb and rupture of the trunk. Injuries and fatalities are reported from Florida, Tanzania and Australia. In Tanzania, there is a belief that these animals attack only when the river in which they live is polluted and the amount of fish and other animals caught by alligators is reduced. In many cases they are too close to human settlements. These animals generally pull their prey into the water and the remains are never seen again. Primate fatality caused by crocodiles has been claimed to have occurred millions of years ago in the Oligocene epoch.

Sea Animals

With nearly 1200 miles of coastline, Florida is an ideal location for people to come in contact with a large variety of marine animals. Swimmers, boaters and divers are all at risk of exposure to offending marine animals. Some of these animals are toxic, such as jellyfish, fire coral, the Portuguese man-of-war, sea anemones, sea urchins, sting rays and some fish. While these animals can and do cause intense pain from envenomation, death is unlikely to occur from the sting unless the individual is allergic to the venom. The intense symptoms caused by the envenomation may lead to the victim drowning.

Other sea animals, such as sharks, barracudas and moray eels, can and do cause severe injury and sometimes death. Sharks, common residents of Florida's warm waters, commit the most common animal bites.

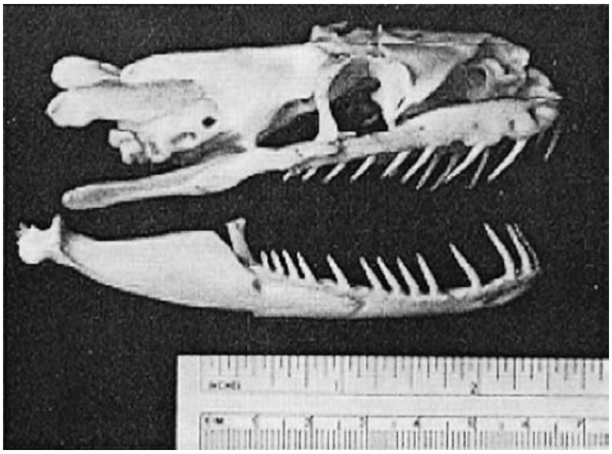


Figure 6 Lateral view of snake skull showing recurved teeth for securing prey. (From McCarty *et al*., 1989.)

They inhabit the same areas that people enjoy. They are found in all tropical and subtropical areas of all oceans, both inshore and offshore and around oceanic islands. They are also frequent estuaries, bays, lagoons and reefs. Their attacks may be provoked, unprovoked or acts of scavenging. Causes of provoked attacks are usually in the nature of divers molesting the animal or a fisherman removing a hook. Unprovoked attacks are of three types. The most common, the hit-and-run attack, occurs in about 80% of cases. These kinds of attack are believed to be for the purpose of investigation. They are termed hit-and-run because they are not usually repeated. The injuries from these kinds of attack are less severe. Surfers, swimmers and waders are most at risk. The hit-and-run attacks are perpetrated mostly by small sharks, 1–2 m, in length and usually very close to shore.

The second type of attack is the 'sneak' and 'bump-and-bite'. These attacks occur in deeper waters, and are generally carried out by larger sharks. The sneak attack is the kind observed in the motion picture *Jaws*, where the victim has no advance warning and the attack is violent. The bump-and-bite attack is the kind in which the shark first circles the victim, actually bumping into it prior to a violent attack. The sneak and bump-and-bite attacks are considered to be intentional and represent aggressive feeding behaviors. Historically, 17 deaths have occurred in Florida waters since 1988. All have been attributed to sneak and bump-and-bite attacks. Sharks possess many rows of very sharp teeth. They are capable of exerting tremendous biting force that can result in the complete avulsion of soft tissues from the bone, neurovascular injuries and destruction of bone. Sharks cannot chew their prey but their manner of feeding allows them very effectively to strip the tissue from the prey, leaving characteristic gouge marks beginning at the site where the shark first grabs the victim and extending down its length (**Figs 7** and **8**). Depending on the size of the victim and of the attacking shark, it is possible for the animal to amputate an appendage in a single attack.

Two other marine animals are barracudas and moray eels. Barracudas are known for their viciousness. This fish can reach lengths of up to 180 cm and weigh more than 50 kg. Unlike sharks, which must bite in a grab and sawing motion (number 8 motion), barracudas can cleanly shear off portions of their prey in one fast attack. These fish have been known to completely remove a hand from a fisherman with one cleaver-like bite. Bites by moray eels are rare. When they occur it is usually the result of provocation, such as an angler removing a hook. The wounds are generally ragged because of the eels' multiple rows of sharp teeth and their movements following attachment.

Discussion

There has long been an anthropological interest in assessing trauma caused by animals. Human fatalities due to animal attacks are not as common as other causes of death. In most cases attacks are not considered forensically relevant unless it involves death. Effects of animal attacks on humans can be both injurious and cause fatality. Attacks may be provoked as a result of intentional predation, of scavenging or defense of territory.

The challenge is twofold: one aspect is determination of cause of death from the remains, and the other is differential diagnosis of bone injuries. Zooarchaeologists have spent considerable time developing differential diagnosis criteria to identify the predator. There have been experiments assessing the injuries on bones caused by many animals (**Table 3**). Damage to bone from these animals can be described as one of four basic types. The first type of mark is the puncture, caused simply by the bone collapsing under the

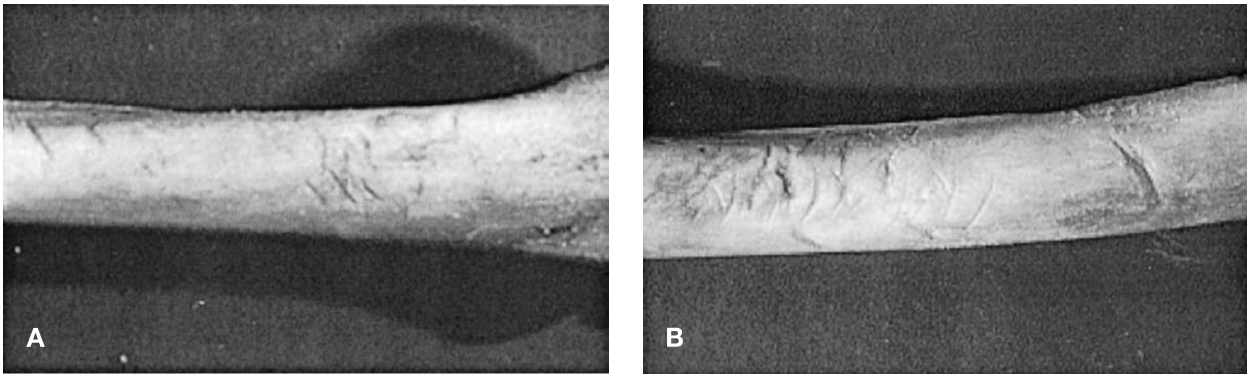


Figure 7 Bite marks from tiger shark on **(A)** tibia and **(B)** femur (From İşcan and McCabe, 1995.)

Figure 8 Bite marks from tiger shark on the proximal end of femur. (From İşcan and McCabe, 1995.)

pressure of the tooth, usually a canine, and leaving a clear imprint of the tooth. The second type is the pit. The animal begins with the softer end and moves toward harder bone. Pits are areas in the bone that did not collapse under gnawing action. The third type, furrow, is scoring that occurs when teeth are moved across relatively compact bone. The fourth type is that of the crushing injury. The bone is crushed and splintered. This type of damage may be caused either by the posterior teeth of carnivores or by the hooves of ungulates.

Archaeological remains show many aspects of animal activity, as illustrated by the groves and other marks left on bones. Gnawing marks are relatively simple to assess but if bones have been crushed by a large animal, such as an African lion or Asian tiger, then the marks require special observation. Mammalian bites, usually from domestic animals or humans, account for a small percentage of visits to hospital emergency rooms and physicians' offices. Wild animal bites are even rarer. The literature contains few accounts of wild mammalian bites and a few reports of surveillance data.

Animal ecology and predator behaviors are thus important areas one needs to understand. There is new research demonstrating wound morphology caused by the animals mentioned above. It is important that a crime scene report be properly prepared before any conclusion is reached as to what the cause of death of a victim was. Individuals and professionals should be consulted as to what animals native to the area may cause wounds, and fatality, of the type observed in the remains.

See also: **Anthropology**: Excavation and Retrieval of Forensic Remains; Skeletal Trauma; Overview; Archaeology. **Causes of Death**: Postmortem Changes. **Entomology**.

Further Reading

Aronson H and Tough SC (1993) Horse-related fatalities in the Province of Alberta, 1975–1990. *American Journal of Forensic Medicine and Pathology* **14**:28–30.

Binford LR (1981) *Bones: Ancient Men and Modern Myths*. Orlando: Academic Press.

Burgess GH, Callahan MT and Howard RJ (1997) Sharks, alligators, barracudas and other biting animals in Florida waters. *Journal of the Florida Medical Association* **84**:428–432.

Cohle SD, Harlan CW and Harlan G (1990) Fatal big cat attacks. *American Journal of Forensic Medicine and Pathology* **11**:208–212.

Table 3 Bone defects caused by various animals

Animal	*General effect*	*Bone involvement*
Dog, fox, wolf	Splintering	Cancellous bones (long-bone ends, ribs, mandibular ramus, scapula, vertebrae)
Rodent	Parallel 'chisel marks' due to gnawing	Compact and spongy bones
Black bear	May scatter bones, but is generally vegetarian	No identifiable damage
Raccoon	Scatters small bones	Damage nonspecific in regards to bones
Deer, sheep	Hold bones with back teeth; may be partially broken to give 'fork-shaped' or 'double-pronged' fragment	Long bones
Pig	Crushing, no cutting or splintering; long bones dragged but not damaged	Small-bone ends, ribs
Vulture	Scatter small bones, may be disposed of	Do not inflict physical damage to bones *per se*
Crow, hawk	Scatter bones and may cause their loss	Nonspecific damage to bones
Crab, fish, turtle	First two may leave shallow marks on bones; all scatter bones in water and on land; land crabs scatter bones, and claws may crush spongy bones slightly	Articular ends, ribs and other spongy bones

* Assembled from Morse (1983).

Haglund WD (1992) Contribution of rodents to postmortem artifacts of bone and soft tissue. *Journal of Forensic Sciences* **37**:1459–1465.

İşcan MY and McCabe BQ (1995) Analysis of human remains recovered from a shark. *Forensic Science International* **72**:15–23.

Kenevan RJ, Rich JD, Gottlieb V and Muench AG (1985) A dog bite injury with involvement of cranial content: case report. *Military Medicine* **150**:502–504.

Kneafsey B and Condon KC (1995) Severe dog-bite injuries introducing the concept of pack attack: a literature review and seven case reports. *Injury* **26**:37–41.

Krogman WM and İşcan MY (1986) *The Human Skeleton in Forensic Medicine*. Springfield, IL: Thomas.

Lauridson JR and Myers L (1993) Evaluation of fatal dog bites: the view of the medical examiner and animal behaviorist. *Journal of Forensic Sciences* **38**:726–731.

McCarty VO, Cox RA and Haglund B (1989) Death caused by a constricting snake – on infant death. *Journal of Forensic Sciences* **34**:239–243.

Morse D (1983) The skeletal pathology of trauma. In: Morse D, Duncan J and Stoutamire J (eds) *Handbook of Forensic Archaeology and Anthropology*, pp. 145–185. Tallahassee: Bill's Book Store.

Opeskin K and Lee KA (1992) Death by greyhound. *Forensic Science International* **56**:51–57.

Rabl W and Auer M (1992) Unusual death of a farmer. *American Journal of Forensic Medicine and Pathology* **13**:238–242.

Roll P and Rous F (1991) Injuries by chicken bills: characteristic wound morphology. *Forensic* **52**:25–30.

Rollins CE and Spencer DE (1995) A fatality and the American mountain lion: bite mark analysis and profile of the offending lion. *Journal of Forensic Sciences* **40**:486–489.

Ropohl D, Scheithauer R and Pollak S (1995) Postmortem injuries inflicted by domestic golden hamster: morphological aspects and evidence by DNA typing. *Forensic Science International* **72**:81–90.

Rossi ML, Shahrom AW, Chapman RC and Vanezis P (1994) Postmortem injuries by indoor pets. *American Journal of Forensic Medicine and Pathology* **15**:105–109.

Sacks JJ, Lockwood R, Hornreich J and Sattin RW (1996) Fatal dog attacks, 1989–1994. *Pediatrics* **97**:891–895.

Tough SC and Butt JC (1993) A review of fatal bear maulings in Alberta, Canada. *American Journal of Forensic Medicine and Pathology* **14**:22–27.

Tuggle DW, Taylor DV and Stevens RJ (1993) Dog bites in children. *Journal of Pediatric Surgery* **28**:912–914.

Wiens MB and Harrison PB (1996) Big cat attack: a case study. *Journal of Trauma* **40**:829–831.

Willey P and Snyder LM (1989) Canid modification of human remains: implications for time-since-death estimations. *Journal of Forensic Sciences* **34**:894–901.

Wright JC (1991) Canine aggression toward people: bite scenarios and prevention. *Veterinary Clinics* **21**:299–314.

Archaeology

J R Hunter, University of Birmingham, Birmingham, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0417

Background

Forensic archeology first emerged in the later 1970s as part of a broader recognition of the role of physical anthropology in medicolegal matters. In Britain where archeology was traditionally seen as a free-standing discipline with roots embedded more strongly in excavation and field skills rather than in the study of human skeletal remains, forensic archeology developed somewhat later. Furthermore, differences in medicolegal practice at scenes of crime between the two countries also serve to confine the role of forensic archeology in Britain to one more specifically concerned with field activities than with postmortem study. Elsewhere forensic archeology plays a less prominent role in criminal investigation, although with formal interest now developing in other parts of northern Europe, South Africa and Australasia. Additionally, the excavation of mass civil or war graves in places such as Argentina, Ruanda and Bosnia have widened the application of archaeological recovery techniques to other parts of the world.

Archeological input to criminal investigation reflects the skill base of modern archeology and its role in society. In Europe and North America archeologists are now extensively employed on behalf of central and regional government in order to identify and assess archeological remains ahead of planning and redevelopment schemes. The archeologist's interest is almost exclusively concerned with understanding about the past from buried material remains; this can be expressed in general terms of finding evidence, recovering evidence in a manner which will maximize its value, and in interrogating evidence in order to reconstruct sequences of past events. As a process this has much in common with law enforcement detection and is now widely recognized as having a role in modern forensic investigation when buried remains are encountered. This skill base provides a collective expertise in landscape analysis, knowledge of resources (photographic and cartographic), geology, land use, site location methods and excavation techniques.

In addition, archeology is a process which has access to a range of associated areas of specialties many of which also have forensic application. These include, for example, aerial photographic interpretation, the identification of pollen assemblages, soil

analysis, animal/human bone differentiation, and the recognition of cremated human materials. Archeologists have also developed techniques which are peculiarly suited to forensic purposes, for example geophysical survey for the detection of shallow subsurface remains. Archeology, like forensic science, is also concerned with scientific analysis across a broad range of material types.

The main areas of interest common to both archeology and criminal investigation might be listed thus:

- Skeletal analysis (physical anthropology)
- Scientific analysis
- Field search
- Excavation and recovery

The first two of these are discussed more fully elsewhere in this volume and are not detailed here other than for crossreference purposes. Skeletal analysis, for example, is a major section in its own right with a recognized modern history in criminal matters as well as in military and disaster scenarios in the United States. Equally, the application of both physical and life sciences is now a fundamental part of modern archaeology and much analytical science is dedicated to matters of provenance or dating. Provenance – the ascribing of a finished product (e.g. pottery) to the location of a natural source (e.g. of clay) – is a well-attested method of evaluating the trade, transport and social economics of earlier societies. Like forensic science and the Locard principle, it ascribes a material to a parent source. The methods used are also common to forensic science, as indeed are many of the materials under study (e.g. silicates, soils, metals, pigments etc.). There are many crossreference points throughout this volume although interplay between the two disciplines of forensic science and archeological science respectively is curiously exclusive.

Absolute dating methods also have some relevance here. Although archeological and forensic time-scales have little in common, radiocarbon dating has some application in demonstrating whether human remains pre- or postdate 1952 (the era of atomic weapon testing which generated high atmospheric radiation levels). This is particularly important in providing coarse dating in instances of stray or displaced human bone. More relevant perhaps is the observation of taphomomic (decay) processes in buried environments and the determination of interval since deposition. There is a growing area of interest in both human and associated modern materials buried over short time-scales.

The number of homicide victims who are ultimately buried to avoid detection is relatively small, perhaps little more than 2% of the total victims, although several of these have been multiple or serial killings and have been of high profile. A greater percentage involves concealment in one form or another. Recent history has emphasized a number of cases which have demonstrated a potential role for archaeology: in the United States these have been highlighted by one of the pioneers of forensic archaeology, Dan Morse, from cases in Florida, Illinois and Texas, and in Britain notably from the Moors Murder enquiry in the early 1960s and the later investigation of the north London garden of Dennis Nilsen in 1984. All these situations were effectively archaeological scenarios which would have been approached and handled in very different ways if taken on a straight archeological basis. However, since the late 1980s, but with some notable setbacks, such cases have been fewer and awareness of the potential value of using archeological techniques has increased among law-enforcement agencies. It is, nonetheless, not simply a straightforward matter of creating awareness or of applying appropriate techniques. The process is two-way: archaeologists also need to recognize the protocols and processes of crime scene work, the time-scales involved and, most importantly, the judicial constraints which surround forensic work. The philosophy and practise of forensic archaeology is very different from that of its parent. Also different are the objectives and the means to achieving those objectives (see below).

Archeology is based on the principles of stratigraphy, namely that the ground subsurface comprises discrete layers of earth which differ in texture, color and physical properties reflecting the effects of both natural and human activity over time. These layers can be generated by natural processes of soil formation or, in areas of habitation, by building, farming and general development caused by a wide range of activities. In rural areas these layers can be of relatively shallow depth above bedrock (i.e. a few centimeters), or in urban areas up to several meters deep reflecting centuries of activity and use. Archeologists try and interpret what each layer represented in past time, and also how the various layers relate to each other chronologically. The order of their formation can usually be determined quite easily. Many modern layers, particularly those relating to building activity, may be able to be dated by modern records.

In effect, the profile of layers in a given area of the subsurface makes a type of statement about the history of that place. It also presents a 'snapshot' in time, in that all the layers have a broad chronological relationship in which one layer which overlies another is clearly of later date (**Fig. 1**). If a grave is dug into the ground the existing layers in that place become disturbed, a relationship occurs between these layers and the grave, and the time-frame of

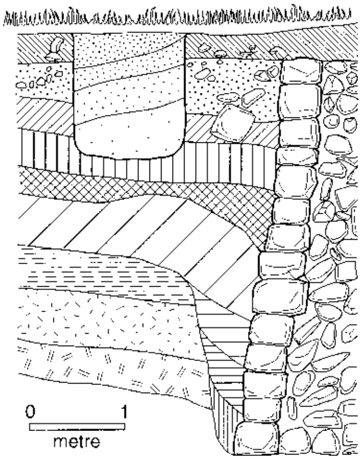


Figure 1 Example of local stratigraphy showing development of layers through time, including imposition of wall (right) and later pit or grave (top left). Drawing by H Buglass.

the overall stratigraphy becomes extended. Archeology involves identifying and interpreting changes of this type.

A grave identified within these layers will have a contextual integrity, no matter how few layers are involved. Recovering or excavating a victim without awareness of related layers (i.e. the layers which are earlier than the grave, the layers which are later than the grave, and the grave itself) will stand to lose much of the evidence that may be available. Questions which the pathologist or medical examiner will try to answer pertaining to interval since death, identity, or the cause/manner of death may be more readily resolved if the edge and sides of the grave can be defined in three dimensions; this effectively identifies the physical boundaries within which the human remains and associated materials will be found. The presence of the associated layers may assist in dating the event.

Human remains often appear during building or development work when bones become unearthed during machine excavation. In such instances it is particularly important that remains are retained *in situ* in order that the relationship of the grave within its buried environment can be retained as much as possible.

Search

Locating buried materials can involve a range of target types, although the most common are human remains. Other targets can include firearms, drugs or stolen goods covering all shapes, sizes and materials of organic, inorganic, ferrous or nonferrous type. In nearly all instances the method of search involves the identification of the disturbance caused by the burial in the first instance, rather than the item buried. Even with homicide victims (for which this entry is primarily intended) the target characteristics can vary according to stature, age, clothing or dismemberment.

The initial phase of search usually involves targeting the most likely locations in a given area using factors of feasibility (i.e. geological suitability, soil cover, land use etc.) in combination with a suspect's movement, psychological profiling and other intelligence. 'Dump-site' analysis as it is known, has been the subject of valuable research in both Britain and the United States: it enables factors such as the relationship between victim and suspect; location of last sighting, or likely distance traveled, to be fed into the search equation. The use of geological maps, existing aerial photographs (vertical and oblique) can facilitate this. Often the process can be carried out as a 'desktop' study and the target areas located accordingly. This is highly desirable as it does not arouse suspicion, run the risk of damaging evidence on the ground, or necessarily incur great expense.

Once the area can be narrowed down, a more detailed approach can be taken. When a victim, or for that matter any object, is buried, the subsurface is disturbed in relation to the surrounding environment. This disturbance can have a number of effects on (a) the subsequent ground vegetation, (b) the immediate topography, and (c) the geophysical signature of the disturbed area.

The original digging of a grave is likely to create a looser, damper, more aerobic area of ground when the grave is filled in. This will almost certainly affect the height or density (or even species) of the resulting vegetation on the ground surface (**Fig. 2**) – an effect that may become more pronounced as the human decay process provides nutrients to an already moister soil medium. The converse might apply if the grave was filled in with stones and the vegetational growth inhibited accordingly. These changes will have a long-term effect. There will also be a shorter term effect on the surrounding ground surface

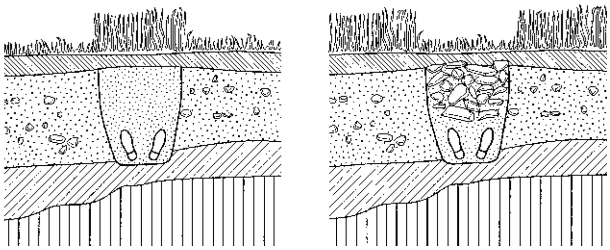


Figure 2 Some potential effects of burial on surface vegetation. Drawing by H. Buglass.

where the excavated spoil was dumped when the grave was first dug. There may also be excess soil providing a low mound over the grave which in time may sink to form a depression as the grave infill consolidates. A secondary depression may occur as the body cavity collapses, and cracking may take place at the grave edges during dry weather. Additionally, the infilling of the grave may contain obvious traces of deep disturbance such as soil or bedrock of different color (e.g. clay or chalk) brought to the surface in an otherwise undisturbed area.

Although many of these features of vegetational, topographic or soil change might be seen at ground level, they are often at their optimum when viewed from the air, some features being particularly prominent through the use of shadows at appropriate times of the day. Aerial reconnaissance is an ideal nonintrusive primary method of search, but its effectiveness may be seasonally specific given that light and land use are variable components. Any photography, however, may be supported by imagery analysis for which there is now highly developed military expertise.

There are a number of methods which can be used to detect buried remains; some of the most commonly used are listed below on a scale which runs from the noninvasive (i.e. those least likely to destroy buried evidence) to the invasive (i.e. those most likely to destroy buried evidence). The list is not comprehensive and is presented only as a general guide; it also tends to reflect the selective narrowing down of the target area. The deployment of these techniques, and their use in sequence (which is strongly advised), depends on a range of factors including the local environment, the geology, the nature of the target, and the interval since burial. Some techniques, particularly geophysical methods, are only suited to certain types of environment or situation (e.g. indoors, outdoors, concrete, soil, clay, etc.) and are wholly ineffective in other environments. It is not possible to outline the complex range of variables in the space available here; search is a specialist area of expertise for which advice should be sought (see below). The primary use of one method may completely negate the effectiveness of another, or may be wholly wasteful of resources. A routine list of techniques might be as follows.

- Aerial photography and reconnaissance for the detection of disturbances identified from the vegetation or topography (shadow or color change) or by soil marks (color).
- Field craft for the detection of discrete areas of vegetational, topographic or soil difference, also for the presence of ground cracking and for the identification of burial locations by using prominent local landmarks or features which may have been used as guide points by the offender (sometimes known as 'winthroping').
- Body-scent dogs for the identification of buried remains by virtue of characteristic odor which may be released from the ground by systematic or selective probing.
- Geophysical survey for the detection of shallow subsurface features, typically by using electrical resistance (to detect disturbance which has affected local moisture content of soils), magnetometry (to detect disturbance caused by changes to a local magnetic field), ground penetrating radar (to detect responses to changes in ground density using an electromagnetic pulse) and metal detector (to detect both ferrous and nonferrous items associated with a buried individual).
- Probing/sampling for the identification of subsurface disturbance, emission of methane, or thermal change.
- Excavation for the identification of soil disturbance (color, texture and physical properties), and changes consistent with a grave (see above).

The effectiveness of each of these methods depends on a range of factors, not least of which is the ability to detect a disturbance or change within an otherwise undisturbed environment. Also to be considered are effects brought about by the decay process of the body itself. This may not only generate heat (and allow the target to be identified using thermal imaging during the decay activity) but it may also affect the geophysical properties of the grave and inhibit or enhance the likelihood of detection depending on the survey technique used. Variables which might accelerate or delay this decay process include factors of depth, climate, soil pH, wrapping or clothing, water content and presence of oxygen. Optimum methods of search rely on knowledge of as many of these variables as possible.

In Britain a central search advice facility is available under the auspices of the Forensic Search Advisory Group. The group consists of specialists who adhere to a code of practice and who have experience in the search and recovery of buried remains. In the United States the group NecroSearch operates a similar, larger, facility based in the Colorado area. Both groups operate a number of associated research programs on decay and detection and maintain test sites.

Recovery

Although recovering buried remains can occur in a variety of different situations and environments the

process of recovery follows a well-established routine in order to maximize the evidence available. This is based on the awareness that archaeology is a destructive process and that each recovery operation is seen as a nonrepeatable exercise.

When the burial of a victim takes place, three main activities are undertaken: the physical removal of earth from the ground; the deposition of the body in the grave; and the infilling of the grave. Proper forensic investigation of the grave follows this process in reverse order: the grave deposits are removed to see how the grave was infilled; the body is exposed in order to show the manner in which it was disposed; the body is lifted; and subsequently the manner in which the grave was dug is identified.

Once the outline of a disturbance has been identified it is normally half-sectioned (i.e. half of the infill should be excavated), usually by bisecting the long axis at around the midway point leaving an exposed vertical profile of the unexcavated part of the disturbance. This has two main benefits: first, it serves to identify whether the disturbance is in fact a grave or not without excavating (and destroying) the whole disturbance, and secondly, it provides visible evidence in the exposed profile as to how the disturbance was infilled, and how it might most effectively be excavated to resolve questions pertinent to the particular case.

Sometimes, because of depth or other constraints it becomes necessary to extend the working area available by excavating additional space adjacent to the grave itself (having first suitably recorded the grave sides). This allows lateral access to the grave/body and facilitates proper excavation. In situations where this is impossible, it may be necessary to construct struts or planking to wedge across the grave as a working platform. This ensures that the weight of the excavator is taken by the planking rather than the underlying victim (at whatever depth), thus minimizing unnecessary damage to the victim and to any associated materials. Although no two scenes are ever the same, and each has a different situation and timescale, a number of general elements need to be addressed during the recovery process. These are outlined below and should be seen in general terms only, given the great variability of situations and constraints of individual scenes.

Preserving the integrity of the grave It is critical that the integrity of the grave (i.e. the boundaries and area of disturbance) is maintained throughout, in order to preserve the exact situation that occurred when the victim was buried. This not only allows the original scene of crime to be recreated, it also insures the absolute recovery of the individual and any associated materials. Furthermore, determination of the parameters of the grave is also necessary in order to eliminate the possibility of contamination.

Emptying the grave The grave infill is normally removed by trowel in layers which reflect the manner in which the grave was infilled by the perpetrator. This also enables the nature of the grave infill to be identified. If there are no observable layers, the grave infill is excavated in a series of 'spits' typically 10 cm deep in order to provide a controlled removal of the grave deposits.

Objects found within the grave Exhibits discovered within the grave are normally recorded in three dimensions within the layers or 'spits' in which they occur. Recording of this type is now standard practice on archaeological sites and is carried out using an electronic distance measurer (EDM) which is both accurate and rapid. This ensures that individual objects are properly associated with any layers in which they occurred, and that the contents of the grave can be recreated spatially. These exhibits can include items belonging to the individual and his/her clothes, such as buttons, jewellery and the contents of pockets, or items associated with the burial event such as a weapon, papers, clothing and wrapping, or material introduced to the grave infill in an attempt to minimize discovery, such as rubble or lengths of wood.

Recording Because archeology is a destructive exercise the recording process is comprehensive, and wherever possible is undertaken in three dimensions. The outline of the grave is normally planned with reference to permanent base points at ground level in order that it can be relocated in the future; the base and the profile of the grave are planned using graphic conventions such as contours or hachures, and a vertical record of the infill from the half section (above; **Fig. 3**). This section is drawn as part of the basic recording methodology and is essential in demonstrating the dimensional properties of the grave.

On archeological sites human skeletal remains are drawn to scale as part of the recording process; this may not be possible at a scene of crime, but the disposition of the body and the location of any associated features are usually subject to detailed photographic record. The body is normally revealed in its entirety before lifting. The lifting process itself is a delicate operation which takes place bearing in mind that footprints may have survived in the floor of the grave. The method of lifting varies depending on factors of depth, condition and wrapping, but is

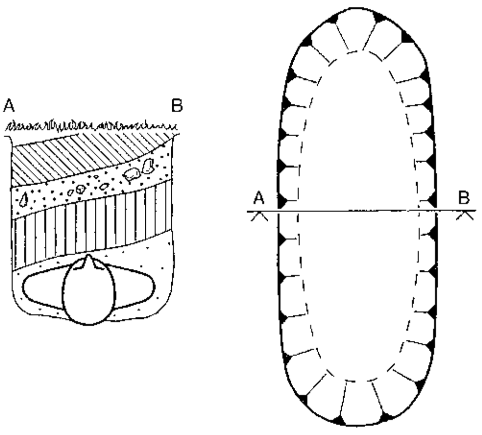


Figure 3 Left: vertical profile of grave infill taken at approximate half-way point across disturbance. Right: hachure plan of overall grave profile on completion of excavation and removal of victim. Cross-reference provided by section line A–B. Drawing by H. Buglass.

often most satisfactorily carried out by sliding a board or metal sheet under the body, although this may have implications for the survival of any foot impressions.

Photography occurs throughout the recovery process, ideally taken from a fixed point at one end of the grave (normally the foot end) and identifies specific rather than arbitary points during the recovery process, for example before excavation, at changes in layers or 'spits' in the grave infill, at the point where significant exhibits occur, when the body is exposed, and when the body has been lifted. These photographs will not only record the main sequences of events but, viewed in reverse order, will recreate the offense in exactly the way it occurred.

Excavation and recording along these lines may be able to answer several important questions about the original crime.

- *How was the grave dug and with what implement?* This may be resolved by examination of the sides of the grave which exhibit mechanical grab marks, shovel (curved edge), spade (straight edge) or pick (short blade) markings. In exceptional instances it may be possible to identify a specific implement which bears characteristic markings or nicks on the blade. This is especially the case in soils which sustain impressions such as heavy clays.
- *Was the grave dug in a hurry or was it carried out carefully in a prepared way?* The professional opinion of the archaeologist may be canvassed in an attempt to determine whether a grave site was premeditated, on the basis of its profile and depth (or even of its primary silted deposits if the grave had been open for some time before burial). This clearly has implications for manslaughter or murder charges.
- *Did the grave contain any evidence of cause or manner of death?* The infill of the grave may bear traces of toxins, blood or firearms discharge and the excavator will be prepared for these within the grave deposits. Toxins will normally be the subject of a sampling strategy by scene-of-crime personnel. The location of bullets/pellets may be determined by the use of a metal detector. Bullets may, for example, have passed through the base of the grave into the subsoils if the victim was shot in the grave itself.
- *Was there transfer of material from offender to grave infill?* Locard's principle points to the natural transfer of material (e.g. fibers, hair, sweat or footwear impressions) into the grave infill. These are often extremely hard to identify or recover, but should be borne in mind during the recovery operation.
- *Was there foreign material in the infill or layer matrices, and if so where did it come from?* During the excavation the descriptions of individual layers or 'spits' are recorded using accepted soil definitions and terminology. Soils or materials which are unlikely to have been dug out from the area of the grave can provide additional evidence. Their provenance may have considerable bearing on convicting the offender.

Conclusion

Forensic archeology is now an acknowledged area of expertise in field search and recovery and has been legally recognized in prosecution and defense arenas during the late 1980s and 1990s in both the United States and Britain, as well as being actively developed in other European countries. It has been recognized as a subdiscipline in its own right and is offered within university degree programs on both sides of the Atlantic. Forensic archeology also features in post-experience and validated postgraduate courses and seminars intended for both archeologists and law enforcement personnel, forensic scientists and associated professionials. Despite its academic origins, its familiarity within sampling, three-dimensional study and recording processes lends itself naturally to scene-of-crime investigation. Furthermore, its extensive repertoire of associated disciplines provides a valuable range of distinctive forensic skills.

See also: **Causes of Death**: Scene of Death. **Crime-scene Investigation and Examination**: Recording;

Collection and Chain of Evidence; Recovery of Human Remains; Packaging; Preservation; Contamination. **Photography and Digital Imaging**: Digital Imaging Enhancement. **Anthropology**: Skeletal Analysis; Morphological Age Estimation; Sex Determination; Determination of Racial Affinity; Excavation and Retrieval of Forensic Remains; Animal Effects on Human Remains; Stature Estimation from the Skeleton.

Further Reading

Boddington A, Garland AN and Janaway RC (eds) (1987) *Death, Decay and Reconstruction: Approaches to Archaeology and Forensic Science*. Manchester: Manchester University Press.

Clark A (1996) *Seeing Beneath the Soil: Prospecting Methods in Archaeology*. London: Batsford.

France DL, Griffin TJ, Swanburg JG *et al*. (1992) A multidisciplinary approach to the detection of clandestine graves. *Journal of Forensic Science* **37**:1435–1750.

Haglund WD and Soirg M (eds) (1997) *Forensic Taphonomy: the Postmortem Fate of Human Remains*. Boca Raton: CRC Press.

Hunter JR (1998) The excavation of modern murder. In: Downes J and Pollard A (eds) *The Loved Body's Corruption*, pp. 197–211. Glasgow: Cruithene Press.

Hunter JR, Roberts CA and Martin A (eds) (1995) *Studies in Crime: an Introduction to Forensic Archaeology*. London: Seaby/Batsford.

İşcan MY (1988) Rise of forensic anthropology. *Yearbook of Physical Anthropology* **31**:203–230.

Killam EW (1990) *The Detection of Human Remains*. Springfield: Charles C. Thomas.

Krogman WM and İşcan MY (1986) *The Human Skeleton in Forensic Medicine*. Springfield: Charles C. Thomas.

Morse D, Dailey RC, Stoutamire J and Duncan J (1984) Forensic archaeology. In: Rathbun TA and Buikstra J (eds) *Human Identification: Case Studies in Forensic Anthropology*, pp. 53–63. Springfield: Charles C. Thomas.

Morse D, Duncan J and Stoutamire J (eds) (1983) *Handbook of Forensic Archaeology and Anthropology*. Tallahassee: Rose Printing.

Rodriguez WC and Bass WM (1983) Insect activity and its relationship to decay rates of human cadavers in East Tennessee. *Journal of Forensic Science* **28**:423–432.

Sigler-Eisenberg BB (1985) Forensic research: expanding the concept of applied archaeology. *American Antiquity* **50**:650–655.

Skinner M and Lazenby RA (1983) *Found! Human Remains*. Simon Fraser University, British Columbia: Archaeology Press.

Snow CC (1982) Forensic anthropology. *Annual Review of Anthropology* **11**:97–131.

Stevens JA (1997) Standard investigatory tools and offender profiling. In: Jackson JL and Bekerian DA (eds) *Offender Profiling: Theory, Research and Practice*, pp. 77–91. Chichester: Wiley.

Assessment of Occupational Stress

K A R Kennedy, Ecology and Evolutionary Biology, Cornell University, Ithaca, NY, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0769

Introduction

The scientific study of bone and tooth modifications produced by habitual activity patterns, when observable in living and skeletal human subjects, holds an accepted place in the protocol of present-day forensic anthropological investigation. Commonly referred to as markers of occupational stress (MOS), these indicators of one or more persistent and long-term activity-induced changes include musculoskeletal stress markers (MSM) and robusticity markers (RM) in circumstances where ligaments or tendons insert onto the cortical tissue of bone via the perosteum, or where there is a hypertrophy of muscular attachments and thickening of the shafts (diaphyses) of bones of the upper and lower extremities. Some indicators of habitual stress do not involve muscular activity, e.g. the formation of grooves along the enamel occlusal margins of incisor teeth among tailors who hold metal pins in the mouth when sewing. Hair and nails may also reflect an individual's life history of habitual activities by development of structural irregularities. Thus the term MOS applies to a variety of anatomical modifications and agents. Physiological and cytological processes pertaining to bone remodeling and the biomechanical affects of stress on bone are well documented, particularly in the medical and anatomical literature of sports medicine, industrial medicine, military medicine and orthopedics.

The naming of specific anatomical modifications of skeletal, dental and soft tissue structures as 'markers of occupational stress' is attributed to Francesco Ronchese of the Boston University School of Medicine whose text, published in 1948, is entitled *Occupational Marks and Other Physical Signs: A Guide to Personal Identification*. However, the diagnoses of activity-induced stress began in the Middle Ages with Agricola's writings on mining hazards and were developed with the advent of the industrial revolution with medical treatises about relief of symptoms suffered by the working classes in farms, factories and trade centers. Anthropologists became interested in MOS from the examination of fossil hominid skeletal remains whereby identifications of squatting facets in Neanderthals and severe occlusal wear patterns of the dentitions of most prehistoric hunting–foraging

populations shed light on the behavioral features of extinct populations. The anthropologist who contributed significantly to the study of MOS was the late J. Lawrence Angel of the National Museum of Natural History, Smithsonian Institution, whose investigations were often associated with cases of personal identification brought to him by the Federal Bureau of Investigation (FBI) and other medicolegal agencies. Angel also applied his MOS studies to skeletal series from ancient Aegean and native American populations.

Bone Responses to Stress

In 1892 Julius Wolff, a German anatomist, published his observation that bone tissue has the capacity to increase or decrease its mass in response to mechanical forces. According to Wolff's Law of Transformation, bone remodeling occurs in well-vasculated subchondrial loci when stress is exerted. The 'functional pressure' or load is dissipated by expansion of the bony framework whereby lipping, spurring, exostoses and thickening of osseous tissue take place. Severe and prolonged stress is observable macroscopically and may be attributed to patterns of habitual and prolonged activity during the life of the individual. Bone resorption occurs in response to persistent compression whereby elevated tubercles and crests are formed by the pull of muscles at the loci of their attachments. However, pull may be related to cortical recession (endosteal growth) when tension results in resorption rather than deposition of bone tissue. Bone remodeling involves both resorptive and depositional functions on periosteal and endosteal surfaces.

Analyses of MOS involve determination of the roles of stress (resistance of bone tissue to the biomechanical action of an external compressional force applied to it), strain (bone deformation or distortion due to tensile or compressive force), shear (a consequence of applied force causing two contiguous portions of a bone to slide relatively to each other in a direction parallel to their plane of contact), and torsion (twisting). These are among the responses of bone tissue to the push and pull of forces exceeding a bone's elastic limits so that loci of stress do not return to their original form. Excessive stress, strain, shear and torsion may result in destruction (necrosis) of osseous tissue.

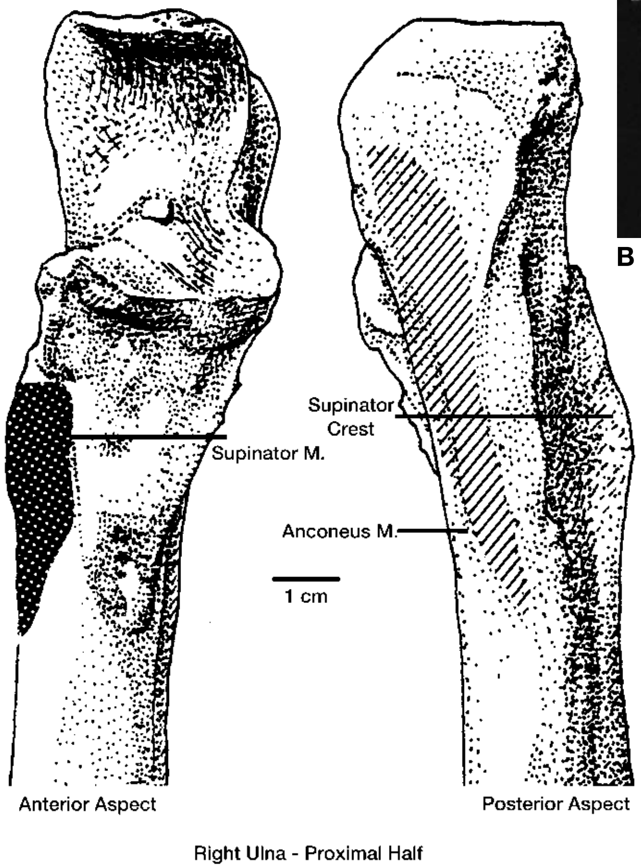
Examples of pathological conditions of osteogenesis and physiology of bone tissue, mechanical adaptations of bones and joints, muscular development and function in relation to skeletal morphology and skeletal plasticity are widely documented in the literature. However, there are fewer references to MOS and their significance in reconstructing the lifestyles of extinct populations (when a skeletal record is preserved) or contributing to a positive identification of an individual skeleton. Interest in these latter issues emerged within the last decades of the twentieth century among forensic anthropologists engaged in studies of ancient populations from archeological sites (paleodemography, paleopathology) and as consultants to medicolegal agencies.

Diversity of Markers of Occupational Stress

Recent studies of the causes of bone remodeling at the cellular and molecular levels have led to revision of Wolff's Law and a more cautious approach to the establishment of direct causality for any specific marker. Factors other than direct response to mechanical loading may be operative, as indicated in the following classification of MOS.

1. The kinds of stressors capable of producing hypertrophy of tendinous or ligamentous periosteal attachments on skeletal organs (enthesopathies and syndesmoses respectively) form irregularities and osteophytes which may be induced by mechanical strain from forces external to the body, as with fractures of the spinous processes of certical vertebrae when heavy loads are carried on the head. Other enthesopathic lesions are induced by internal forces, such as hypertrophy of the supinator crest of the ulna as a consequence of habitual supination and hyperextension of the arm in spear-throwing, ball pitching and use of a slingshot (see **Fig. 2**).
2. Abrasian occurs when bones are in direct contact at joint surfaces, as is the case in severe osteoarthritis when intervening structures deteriorate. The joint surfaces exhibit a polished appearance at areas of contact (eburnation).
3. Attrition is a condition of wear on enamel, dentine and other dental structures resulting from a variety of habitual behaviors, e.g. tooth grinding, ingestion of abrasive particles from ambient dust or incorporation of gritty substances in food preparation processes, as well as use of the mouth as an accessory hand when objects are held between the teeth and moved about in the mouth (see **Fig. 1**).
4. Trauma inflicted on bone and dental tissues may be due to sudden or gradually imposed stress. In cases of bone fracture, osteogenesis originates at some distance from the line of the break in the periosteum and endosteum, then the hiatus is enveloped with replacement of a fibrocartilagenous callus. As new bone advances into the callus, irregularity of the area marks it as the original site of trauma. Dental trauma occurs with evulsion, chipping of enamel and displacement of teeth as a consequence of abrupt or gradual stress.

Figure 1 Human maxilla with complete permanent dentition showing severe wear of the anterior teeth. This is attributed to the habitual practice of using the mouth to hold oblects. Specimen is from a prehistoric archaeological site at Mahadaha, India. Courtesy of Dr J R Lukacs and the University of Allahabad.



A

B

Figure 2 (**A**) Diagrammatic representation of the proximal end of a human right ulna showing hypertrophy of the supinator and anconeous muscle attachments. This may be a consequence of habitual supination and hyperextension of the fore-arm in spear throwing, ball pitching and use of a slingshot.
(**B**) Hypertrophy of the supinator crest in two human left ulnae from a prehistoric archaeological site at Sarai Nahar Rai, India.

5. Bone degeneration is evident at joints and atrophic patches on adjacent bone surfaces due to loss of volume from osteoporosis, conditions associated with the aging process which are accelerated by obesity, heavy physical labor and a broad spectrum of stressors related to lifestyle. Osteoarthritis is a form of chronic arthropathy characterized by eburnation, cysts, formation of osteophytes, pitting and sclerosis. There is a thin line distinguishing arthritic diseases with an inflammatory condition of degenerative joint disease, traumatic arthritis and osteoarthritis deformans in which little or no soft tissue is involved.
6. Nutritional deficiency not only compromises an individual's attainment of full ontogenetic development, as reflected in skeletal maturity, growth, body size and stature, but may be attributed to specific irregularities or bone shape (platymeria, platycnemia, platybrachia) and degree of curvature (bowing). The latter condition may not be related to habitual activity patterns in every case since it is related to pathological conditions of treponemal diseases and rickets. Defects of dental enamel, including hypoplasial lines and pits, preserve a record of nutritional stress during the early years of an individual's life. Analogous markers of episodes of interrupted development are observable radiographically on long bones (Harris lines).
7. In a majority of skeletal samples from different prehistoric and modern cultures, greater muscular–skeletal robusticity is encountered in males than in females, a feature attributed to the harder physical labor in which men engage. Apart from size differences of bones, which may be under the control of genetic and hormonal factors, males, in general, exhibit more robust loci of muscular insertions. Another indicator of sexual difference is the higher frequency of spondylosis and spondylolithesis in the vertebrae of male subjects.
8. Racial differences in the incidence of osteoarthritic modifications of knees and pelves between Asian and European populations have been reported, but studies of MOS in ancient and modern populations suggest that cultural differences are more significant than features attributed to ancestry and inheritance. Thus the high incidence of squatting facets on the tibiae of South Asian laborers, when compared to their infrequency in working class people of European descent, is more directly related to differences in resting postures than to genetics.
9. Age differences in expressions of MOS are evident in cases of advanced osteoarthritis in older individuals who had engaged in hard physical labor over long periods of time, but it is not unexpected that any particular MOS could be modified over the course of an individual's lifetime. There is the potential for early established markers to become less evident with the passage of time under circumstances of discontinuation of the behavior which induces the feature and the action of bone tissue resorption, but this is a subject for future research (**Table 1**).

Interpretation

As expressions of bone plasticity under pressure from internal and extracorporeal forces, MOS have not been assigned to disorders of metabolism, biochemistry, hormonal and enzymatic imbalances, pathology or neural and vascular disorders. However, certain stress responses, such as those resulting in osteoarthritis and other threats to the integrity of the body, overlap with classifications of disease. And biochemical and hormonal functions have a role to play in the form and function of bone tissue.

Because experimentation is limited to field and laboratory studies of nonhuman mammals, many MOS have been inferred from historical and archaeological sources, ethnographic accounts and clinical records. Hence interpretations vary in determining 'causes' of bone and dental irregularities assumed to be responses to habitual activity patterns. Many modern forensic anthropologists support the view that a marker may be attributed to a wide range of cultural practices (not necessarily to a single activity), and that the skeleton registers a mosaic of activities over the course of an individual's lifetime.

As MOS are interpreted in relationship to the entire skeleton and overall patterns of stress, differences of sex and age become incorporated in the analysis of skeletal remains. Today there is a shift away from the traditional practice of asserting that an irregular morphological feature is the sign of a specific habitual activity towards the keener perception of how bone remodeling takes place, i.e. determining the kinds of modification bone tissue may assume within a broad configuration of behaviors. Accuracy in identification of MOS depends on established standards of observation, availability of well-documented skeletal samples for comparative purposes, classifications of degrees of expression of a real or presumed marker, and other currently accepted approaches to replace the more anecdotal and untestable methods pioneered by earlier investigators. Closer collaboration of forensic anthropologists with biomechanicians, cell and molecular biologists, geneticists, biochemists and physiologists have enhanced recent research efforts in personal identification of human remains and paleodemographic reconstructions of the lifestyles of ancient populations.

Table 1 Examples of markers of occupational stress reported by forensic anthropologists and anatomists

Skeletal component	*Anatomical structure*	*Stress factor*	*Activity*
Acoustic meatus (ear canal) of the temporal bone	Auditory exotosis	Exposure of the ear canal to cold water	Diving to harvest marine foods
Vertebrae	Compression of vertebrae ('Snowmobiler's back')	Vertical compressive force	Sledding and tobogganing over rough terrain; snowmobile riding
	Fracture of cervical vertebrae C-6 and C-7 ('Milker's neck')	Hyperflexion of the cervical bodies	Milking a cow when the milker's head is pushed against the animal's flank and the animal shifts position, thereby moving the milker's neck
Ribs/sternum	Flattening of ribs 6–8 with curvature of the lower sternum	Immobility of the thoracic cage	Wearing corsets with rigid stays which are straight and do not permit movement of the ribs, as worn by women in the eighteenth century
Ulna	Hypertrophy of the supinator crest	Supination of the arm	Spear-throwing, pitching a ball, use of sling and atlatl
Elbow	Lateral epicondylitis of the joint ('dog-walker's elbow')	Sudden tugging and traction on the extended and pronated arm	Walking a dog on a short leash when the animal is not trained to heel
Thumb	Fracture along transverse or longitudinal plane of the diaphysis ('cowboy's thumb')	Fracture	Gripping the saddle horn while flying off the saddle, as in a rodeo, or while riding mechanical bar room bulls
Sacrum	Accessory sacroiliac facets at the level of the second posterior sacral foramina and adjacent to the posterior superior iliac spines	Weight bearing, vertical loading in flexion, axial compression of the vertebral column	Carrying infants or other loads on the back over the lumbar–sacral region
Tibia	Flexion facets at the anterior surface of the distal end of the tibia at the ankle ('squatting facets')	Flexion of the knee	Squatting as a resting posture
Knee	Iliotibial band of irritation, usually of one leg, leading to osteoarthritic modifications of the joint ('Musher's knee')	Rapid hyperextension of the leg	Sharp backward kicking of the leg executed by a team driver at the rear of a dog sled to spur team to greater speed over icy and snowy terrain
Teeth	Serrated occlusal surfaces of incisors	Wear	Holding and cutting thread by tailors and seamstresses
	Anterior tooth loss	Trauma, wear	Using the teeth for power grasping in holding sled reins or fish lines, or results of wrestling or fighting

See also: **Anthropology**: Overview; Skeletal Analysis; Morphological Age Estimation; Sex Determination; Determination of Racial Affinity; Excavation and Retrieval of Forensic Remains; Bone Pathology and Antemortem Trauma in Forensic Cases; Skeletal Trauma; Stature Estimation from the Skeleton. **Clinical Forensic Medicine**: Overview. **Identification/Individualization**: Overview and Meaning of ID. **Pathology**: Overview.

Further Reading

Angel JL (1996) Early skeletons from Tranquility, California. *Smithsonian Contributions to Anthropology* **2**:1–19.

Capasso L, Wilczak CA and Kennedy KAR (1999) Atlas of occupational markers on human remains. *Journal of Palaeopathology* **3**:1–183.

Enlow DH (1963) *Principles of Bone Remodeling*. Springfield, IL: Charles C Thomas.

Frost HM (1998) Changing concepts in skeletal physiology: Wolff's Law, the mechanostatt, and the 'Utah Program'. *American Journal of Human Biology* **10**:599–605.

Kennedy KAR (1983) Morphological variation in ulnar supinator crests and fossae, as identifying markers of occupational stress. *Journal of Forensic Sciences* **28**:871–876.

Kennedy KAR (1989) Skeletal markers of occupational stress. In: İşcan MY and Kennedy KAR (eds) *Recon-*

struction of Life from the Skeleton, pp. 129–160. New York: Alan R Liss.
Martin RB and Burr DB (1989) *Structure, Function and Adaptations of Compact Bone*. New York: Raven Press.
Merbs CF (1983) Patterns of activity-induced pathology in a Canadian Inuit population. *Archaeological Survey of Canada* **119**.
Ronchese F (1948) *Occupational Markers and Other Physical Signs: A Guide to Personal Identification*. New York: Grune and Stratton.
Wilczak CA and Kennedy KAR (1998) Mostly MOS: technical aspects of identification of skeletal markers in forensic anthropology. In: Reichs KJ (ed.) *Forensic Osteology*, 2nd. edn, pp. 461–490. IL: Charles C Thomas.
Wolff J (1892) *Das Gesetz der Transformation der Knochen*. Berlin: A. Hirschwald. (English translation published by Fischer-Verlag, 1986.)

Bone Pathology and Antemortem Trauma in Forensic Cases

M Steyn, Department of Anatomy, University of Pretoria, Pretoria, South Africa
M Y İşcan, Adlı Tip Enstitusu, Cerrahpasa, Istanbul, Turkey

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0759

Introduction

The presence and diagnosis of antemortem pathology on human skeletal remains can sometimes assist in the personal identification of unknown human remains. The disease should have been of such nature and severity that the individual and his close contacts would have known about it. If records of the medical condition, and in particular radiographs are available, a positive identification, provable in court, can be made. This may be of particular use in cases where no other means of individual identification exist, e.g. no dental records.

Not all diseases affect bones. Arthritis, for example, is mainly a disease of the musculoskeletal system, but very few of the 200 known varieties leave traces on bone. For any disease to alter bone takes a long time, and therefore usually only diseases of chronic nature are observable. These markers of disease are also often nonspecific in nature, which precludes a positive diagnosis. One of the problems with proper recording of skeletal disease is the lack of consistent, standardized terminology. Another problem that may arise in diagnosis is the incompleteness of the remains. Often only parts of the skeleton are present, whereas the diagnosis of some diseases may require the presence of smaller bones such as those of the hand or foot. Police and other people involved in this type of work should, therefore, be trained to make a special effort to retrieve all fragments.

Despite these problems, many advances in the study of bone pathology have been made over the last few years. Several books on the subject have been published. Although the main theme of their study was disease in history, the descriptions of osseous changes are also applicable to bones of forensic origin. Traditional pathologists are seldom confronted with dry bones, and therefore much of the groundwork has been done by paleopathologists. This field of study has also gained impetus with the formation of the North American and European Associations of Paleopathology.

When examining lesions on bone, care should be taken to determine whether the observed feature is, in fact, pathological and not part of the normal variation seen in all human populations. Detailed knowledge of the morphology of normal bone is therefore essential. This is often true in younger individuals, where normal porosity in metaphyses, due to growth, may be misinterpreted. When investigating remains for evidence of disease, the observer should look for changes in shape and size, abnormal bone loss or formation, fractures and dislocations. The left and right sides of the body should be compared, in order to detect any asymmetries. The distribution of lesions, for example on what bones they occur and on what surfaces, is very important.

Diseases are generally divided into a number of broad categories, namely congenital (inherited), infectious, traumatic, degenerative, circulatory, metabolic and proliferative (including malignant diseases). Antemortem bony lesions are found in diseases from all these categories. Dental disease falls outside the scope of this discussion, but may be found in many other texts. In order to diagnose any of these diseases, the remains should, of course, have undergone a complete anthropological investigation, and it is assumed that the age, sex and population affinity of the individual are known since this information is often necessary before a diagnosis can be made.

There are a number of nonspecific signs of disease on skeletons, which may indicate poor health, malnutrition or anemia in general. These may help to make a general assessment of health status and living conditions of an individual, although they cannot be used for personal identification.

Congenital Diseases

Many congenital diseases are incompatible with a normal life span, and may have led to early death in prehistoric situations. With the onset of modern medicine, however, some of these conditions are treatable or manageable, and the patients may survive for longer. A case such as that illustrated in **Fig. 1** would have been obviously very recognizable in life. This individual, from a modern skeletal collection, suffered from an *in utero* defect in the migration and closure of the first and second branchial arches which led to incomplete formation of the nose area and hypertelorism (eyes wide apart). Cases like this are extremely rare, more so in countries where modern medical facilities are available.

Some of the more common conditions include spina bifida, hydrocephalus, craniostenosis, and cleft palate. Spina bifida is a nonclosure of the spinal canal, where the two halves of the neural arches of the vertebrae have not fused. This occurs most commonly in the lumbosacral area. It may be asymptomatic if small, and is then called spina bifida occulta (**Fig. 2**). People with spina bifida may have abnormalities of the overlying skin, and sometimes also excessive growth of hair in the area. In severe cases the spinal cord is affected, resulting in neurological disorders such as paraplegia and incontinency.

Hydrocephalus involves an obstruction of the flow of cerebrospinal fluid, resulting in fluid accumulation in the lateral ventricles. Since this condition usually prevails since birth, the patients may have a very large and misshapen brain case, and in the case of young individuals the fontanelles close much later than expected, as was the case of the individual illustrated in **Fig. 3**. This child, although about three years old, as judged by the dentition, still had a wide open anterior fontanelle. Hydrocephalus may also be associated with other diseases of the neurological system, such as meningitis, abscesses and tumors.

Figure 2 Spina bifida occulta (Pretoria skeletal collection). Photo: N. Lizamore.

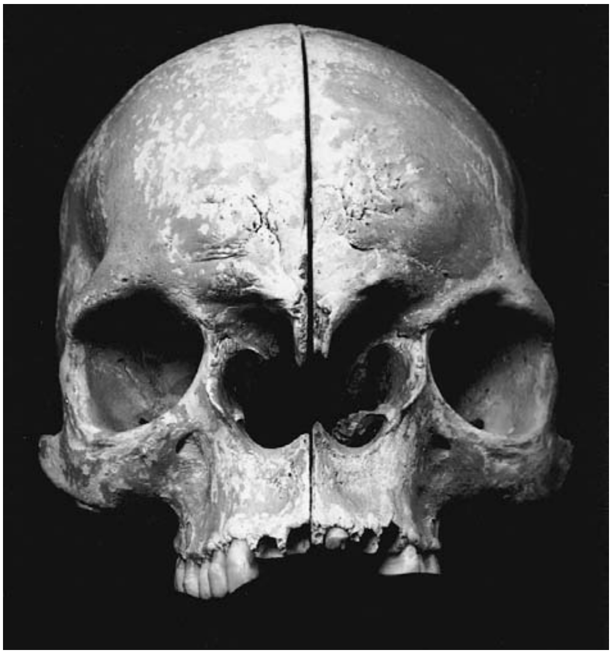


Figure 1 Individual with congenital defects of the face (Pretoria skeletal collection). Photo: P. Birkholz.

In the case of craniostenosis, some of the cranial sutures close prematurely, resulting in abnormal head shapes. Premature closure of the sagittal suture, for example, will lead to an elongated head with a prominent forehead (scaphocephaly), whereas fusion of the coronal suture produces a short skull with a very high forehead. Conditions like these would have been very noticeable during life.

Cleft palate may occur alone or in combination with a cleft lip. A cleft palate is due to incomplete formation of the hard palate, whereas cleft lip also involves the maxilla. It may be complete or incomplete, unilateral or bilateral. In modern remains it can be expected that an attempt at some sort of reconstructive surgery would have been made.

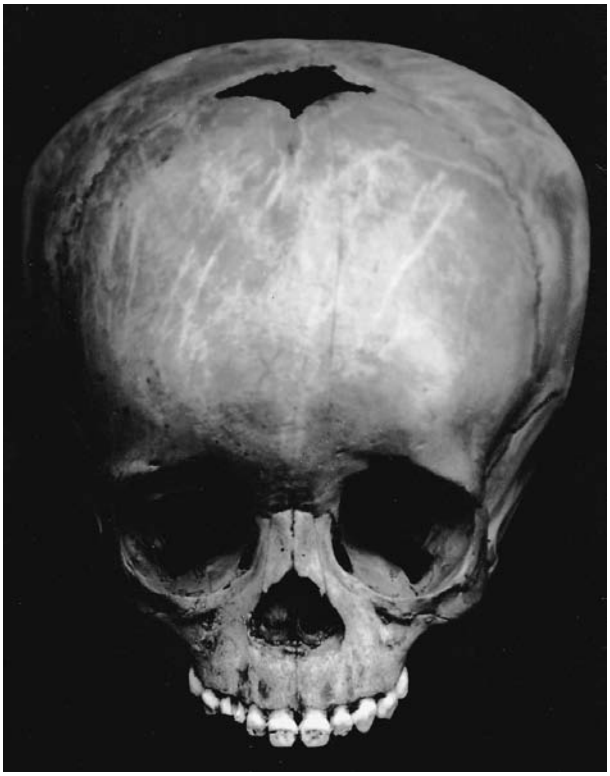


Figure 3 Possible hydrocephalus in a 3-year-old child (Pretoria skeletal collection). Note the large, open fontanelle. Photo: P. Birkholz.

Figure 4 Periostitis in a long bone (Pretoria skeletal collection). Photo: P. Birkholz.

Infectious Diseases

A number of infectious diseases may leave signs on bones, most of which would have been of a chronic nature. It is not possible, from bones alone, to diagnose acute diseases such as gastroenteritis or pneumonia. Advanced chronic infectious disease such as leprosy and syphilis with bone involvement are seldom seen in affluent societies, although they may be more common in less-developed countries. Osteomyelitis, on the other hand, is primarily a bone disease, and sometimes becomes chronic even after treatment with antibiotics, and may therefore be seen in all societies. Acquired immunodeficiency syndrome (AIDS), although of chronic nature, does not affect bones *per se*, but is often associated with other infectious diseases such as tuberculosis, syphilis, and fungal infections of the lungs. In countries where tuberculosis is common, AIDS sufferers often die of tuberculosis even without the AIDS being diagnosed.

Periostitis of bone occurs as a reaction to pathological changes on the underlying bone. It results in new bone formation by the inner layer of the periosteum, and layers of new bone are laid down on the surface of the underlying bone. These bone deposits may be irregular and of varying thickness, often with a woven appearance (**Fig. 4**). It may be due to infection, as in the case of osteomyelitis, or result from trauma or even in the recovery phase of deficiency syndromes (e.g. vitamin C deficiency). Periostitis is usually nonspecific, and occurs in many diseases. It may, however, sometimes lead to a definitive diagnosis. In these cases the distribution of the lesions may give valuable clues, as in the case of syphilis where the lesions are commonly found on bone surfaces close to the skin (e.g. anterior surface of the tibia).

The term osteomyelitis generally describes infection of the bone, caused by any of a number of microorganisms. Most of these infections (80–90%), however, are caused by *Staphylococcus aureus*. Microorganisms may reach the bone (1) through the bloodstream, (2) by the extension of an adjacent infection, or (3) directly via trauma or surgery. The last two are the most common in developed countries, where the jaws and skull are often involved due to dental infections or sinusitis. It is also common in bones of the feet and hands in diabetic patients. Hematogenous osteomyelitis is more common in developing countries, where long bones of children and vertebrae of adults are often involved. The disease usually starts as an acute infection, but may become chronic in about 20% of cases.

Hematogenous osteomyelitis usually starts from the bone marrow, and then penetrates the endosteum.

The bone cortex then becomes infected, and the infection spreads to the periosteum, where it may form a subperiosteal abscess. Penetration of the periosteum may cause sinus tracts (cloacae) through the cortical bone. Small or larger portions of the bone may undergo necrosis, causing a sequestrum (fragment of dead bone). The infection may also spread to adjacent joints, causing septic arthritis. In some cases the infection may become localized and form a chronic area of infection, which is then called a Brodie's abscess. The chronic infection may also stimulate osteoblastic activity, which results in new bone formation under the periosteum (involucrum).

Bones of individuals suffering from osteomyelitis thus show bone destruction, subperiosteal new bone formation, bone deformation, sequestration, and cloacae formation. It may also lead to excessive growth of the affected bone. Osteomyelitis usually leaves bone changes even in well-healed cases. Septic arthritis as a complication of osteomyelitis, is often destructive, and may lead to ankylosis of the bones of the joint.

Tuberculous osteomyelitis is not commonly seen in developed countries, but may occur in compromised patients (e.g. with diabetes or AIDS). In developing countries it usually occurs in young adults and adolescents. It may be slower in onset than is the case in pyogenic osteomyelitis, but is usually much more destructive and difficult to control. It is most common in the spine and long bones where it causes extensive necrosis often resulting in sinuses that drain to the skin. When it occurs in the spine it is called Pott's disease, and here it often leads to compression fractures which in turn give rise to serious deformities such as kyphosis and scoliosis. Next to spinal involvement, the hand and foot bones are commonly affected, followed by ribs, tibia and fibula. Any joint can be diseased, and this is usually associated with infection of the adjacent bones. The knee, hip and elbow joints are most often involved. Skeletal lesions develop in less than 7% of individuals with tuberculosis.

Syphilis (*Treponema pallidum*), although still a common disease in many countries, very seldom progresses to a stage where the bones are involved. There are four syndromes associated with treponemal infection (pinta, yaws, nonvenereal syphilis and venereal syphilis), although they are difficult to distinguish from each other as far as the bone lesions are concerned. Some attempts have been made to do this, mostly by looking at the frequencies of involvement of various bone. Venereal syphilis is either congenital or acquired. In the congenital form the disease is transmitted to the fetus *in utero* via the mother. Severe disease may develop, of which the bone lesions include periostitis and diaphyseal osteomyelitis in young children, and saber-shin deformation of the tibia, gummata on the long bones and deformation of the teeth in older children. In acquired syphilis, the tibia is most frequently involved, followed by various bones of the skull, but virtually all bones can be affected (**Fig. 5**). This usually manifests as periostitis and/or osteitis and gummata. In osteitis the bone itself is affected, often with thickening and an irregular appearance of the whole bone. Due to the thickening of the cortex, the medullary cavities are small. Gummata form as a result of thrombosed blood vessels and toxic products of the microorganism. They vary in size, and may occur in the medulla or subperiosteally. The bone is destroyed in the region of the gummata, but the surrounding bone becomes sclerotic. Skull lesions typically involve erosion of the cranial vault by gummata and nasal–palatal destruction. Skeletal lesions occur in 10–20% of cases with venereal syphilis.

Leprosy is extremely rare in developed countries, but still occurs relatively frequently in poor tropical countries. Various skeletal manifestations that can be found include lepromatous osteomyelitis and periostitis, neurotrophic bone and joint lesions, and ordinary osteomyelitis due to secondary infection. As many as 50% of leprosy patients may have bone lesions.

Figure 5 Syphilitic lesions on a tibia (Pretoria skeletal collection). Left, diseased tibia; right, normal tibia. Note the periostitis and bowing of the affected bone. Photo: P. Birkholz.

Although all bones may be affected, the face and small bones of the hand and feet are most commonly diseased. In the face, the nasal bones, nasal spine and hard palate are often destroyed, and the upper incisors frequently lost (**Fig. 6**). Bone absorption is the most common feature of the hand and foot involvement, and usually starts in the distal phalanges.

Several other infectious diseases, such as brucellosis, parasitic infections and a variety of fungi may leave bone lesions, but they are relatively rare.

Traumatic

Healed traumatic lesions are possibly the most helpful of all antemortem bone changes as far as positive identification is concerned. They leave long-lasting signs, which are usually recorded on radiographs. Mostly the trauma occurs as a single incident, which may be well remembered by friends or relatives. Severe trauma is often followed by surgery, where various devices such as pins and plates are used.

It is useful to include all surgery where these devices are used into this category, although they are of course not all due to traumatic lesions. Prostheses such as hip replacements usually follow after degenerative disease (**Fig. 7**), whereas various forms of metal clips and/or wiring may occur in the sternum after open thoracic surgery. Other devices, such as pacemakers, silicone implants, artificial blood vessels, and heart valves, may also be associated with the remains. Although any fracture may be surgically treated, pins and plates are most commonly found on the femur (**Fig. 8**). In 1988 in the USA it was estimated that about 4.6% of the population had at least one medical device implant. The importance of these devices, depending on their size, is that they can usually be traced back to their manufacturers, and are often issued with a unique serial or production lot number. Manufacturers are supposed to keep records of who these devices are sold to, and in this way it can be traced back to the individual.

Figure 7 Prosthesis of the hip joint (Pretoria skeletal collection). Photo: P. Birkholz.

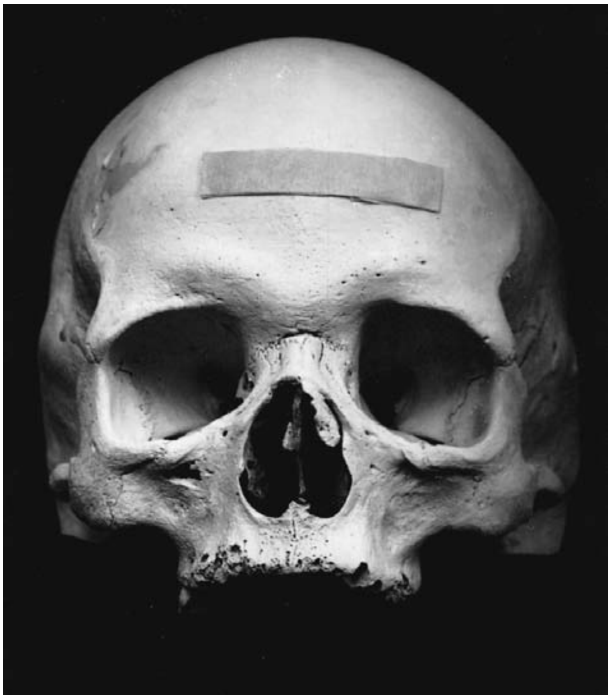


Figure 6 Forensic case with possible leprosy. Note the deformity of the nose, and the resorption of the alveolar bone resulting in antemortem loss of the upper central incisors. Photo: P. Birkholz.

Fractures are usually either pathological, or due to trauma. Pathological fractures occur where underlying disease, such as osteoporosis or carcinoma, is present. In traumatic fractures, the type of fracture sustained depends on the amount and mechanism of force. With direct force the bone may break at the point of impact, usually with much soft tissue damage. Indirect force causes the bone to break some distance from where the trauma occurred, and therefore the soft tissue damage is less. If the direct force is tapping (direct blow), a transverse fracture is caused; if it is crushing, comminuted fractures can be caused. Indirect forces may be twisting, causing spiral fractures; angulating, causing transverse fractures; angulating combined with axial compression, causing

Figure 8 Surgical intervention following a femur fracture. Note the severe shortening of the bone (Pretoria skeletal collection). Photo: P. Birkholz.

Figure 9 Healed fracture of tibia (Pretoria skeletal collection). Photo: P. Birkholz.

a partly transverse fracture with a separate triangular fragment; or a combination of twisting, angulating and axial compression which causes short, oblique fractures. Avulsion fractures occur where muscle action pulls off the bony attachment of the muscle. The above descriptions apply mainly to long bones. Other bones, such as vertebrae, can sustain crush or compression fractures.

Fractures heal by the formation of callus (**Fig. 9**), which undergoes complete or incomplete resorption with time. Complications of fractures which may leave long-lasting signs on bone include incomplete healing with pseudoarthrosis (false joint formation), bone deformity, traumatic arthritis, joint fusion and traumatic myositis ossificans. Amputated limbs are easy to recognize, and may lead to a positive identification (**Fig. 10**). Usually fractures and their complications leave signs long after the event has occurred, thus providing a personal history of the individual.

Face and skull fractures follow after direct trauma. Depressed cranial fractures are mostly due to localized trauma, and may heal with a depressed area still visible on the skull. Healed nasal fractures are rather commonly found (**Fig. 11**), but fractures may be found in any of the facial bones. In severe trauma, such as motor vehicle accidents involving the face, multiple fractures are common, and surgical intervention is needed. Since the mandible is tightly fixed at its two ends, it usually fractures in more than one place.

Dislocation in itself does not leave signs on bones, but may do so if the bone remains dislocated for a significant period of time. In the case shown in **Fig. 12**, a secondary articular facet developed on the scapula. In this position, the arm of the individual would have been dysfunctional, and permanently laterally rotated.

Sharp trauma, such as those caused by knives, axes and pangas, may also leave signs on bones (**Fig. 13**). If the bone tissue around the injury shows signs of healing, the individual survived for some time after the incident. The first signs of healing after bony trauma usually appear after about 7–10 days.

Figure 10 Amputated femur (Pretoria skeletal collection). Photo: P. Birkholz.

Peri- and postmortem trauma can be very difficult to distinguish. Perimortem trauma is usually associated with a 'green bone response' or bending, due to the presence of collagen in the bone. Fresh bones may splinter, forming irregular edges. Dry bone is usually brittle, and postmortem trauma may therefore produce shattering resulting in smaller bone fragments.

Degenerative

Degenerative disease may also be helpful in determining the identity of an individual. It usually occurs in older individuals. Radiographs of the diseased joint often exist, and in some cases surgical procedures, such as implantation of prostheses, may have been done. In a case from South Africa, for example, the body of an elderly individual with severe arthritis of the hip was discovered near a small town. Positive identification was easy when records of a missing person were found who was already booked in for a hip replacement.

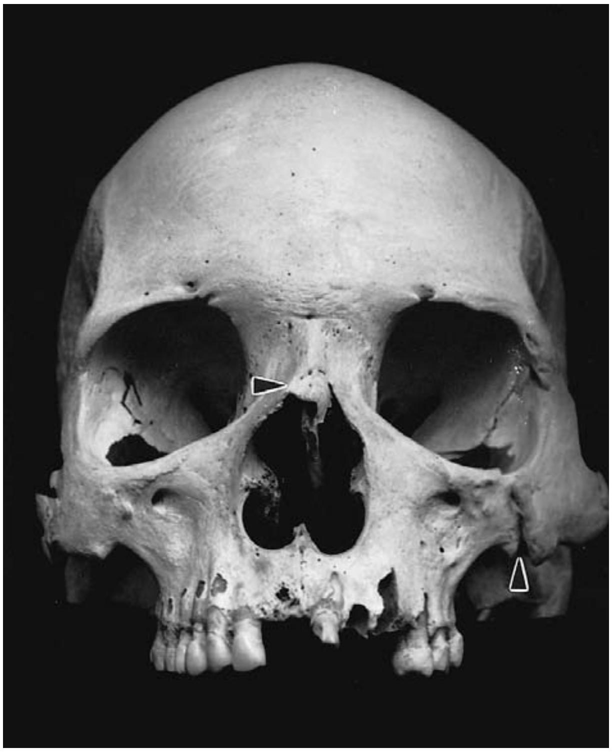


Figure 11 Healed nasal and zygomatic fractures (Pretoria skeletal collection). Photo: P. Birkholz.

Although there are many causes of arthritic disease, only a few can be diagnosed with the help of dry bone only. These are osteoarthritis, vertebral osteophytosis, traumatic arthritis, rheumatoid arthritis, ankylosing spondylitis, infectious arthritis and gout. Of these diseases, only osteoarthritis and vertebral osteophytosis are truly degenerative in nature.

Bone changes due to old age are very common, and can be seen in most people over the age of 50.

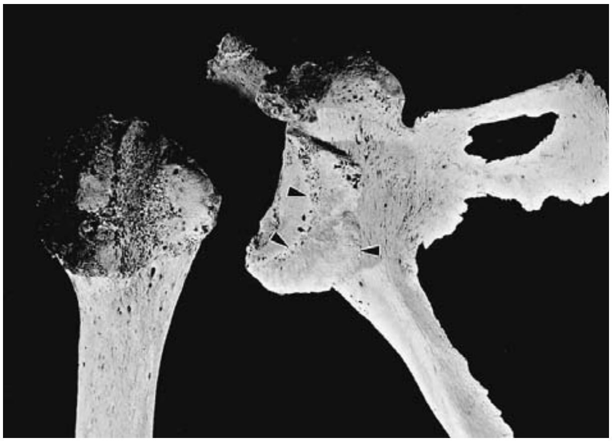


Figure 12 Forensic case with secondary articular facet for head of humerus on scapula. Photo: M. Loots.

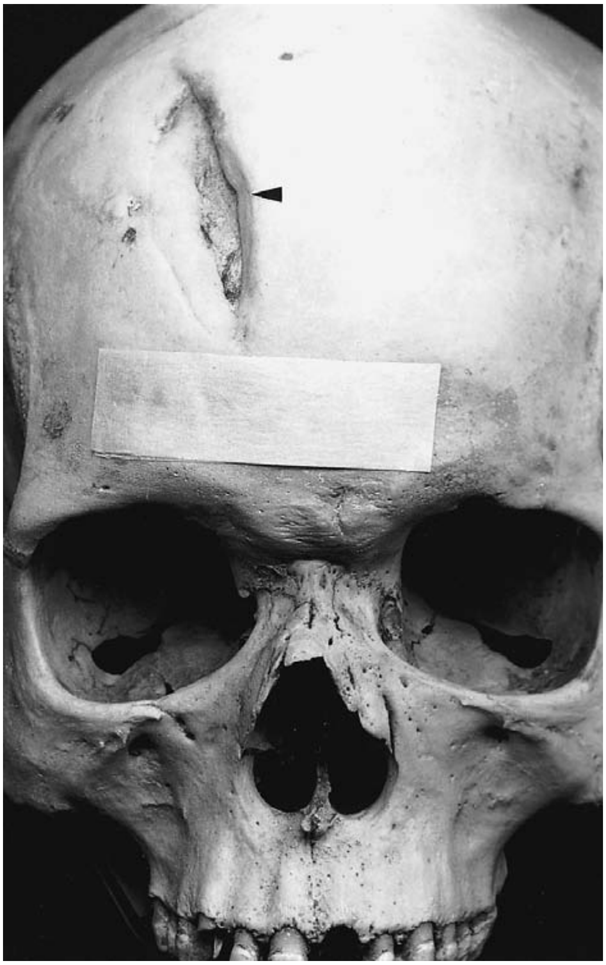


Figure 13 Healed sharp trauma on the skull (Pretoria skeletal collection). Photo: P. Birkholz.

Figure 14 Eburnation of distal femoral joint surface (Pretoria skeletal collection). Photo: P. Birkholz.

Osteoarthritis is usually characterized by deterioration of the joint cartilage and formation of new bone near the joint surfaces. The subchondral bone shows irregular pits, and lipping or osteophytes form near the margins of the joint. If the overlying cartilage completely disappears, the bone is exposed. The bone may then become sclerotic with a polished (eburnated) appearance (**Fig. 14**). Weight-bearing joints, such as the hip and knee, are most commonly affected, whereas distal interphalangeal joints are also frequently affected. Although more than one joint is usually affected, the disease is not as generalized as is the case with rheumatoid arthritis.

Rheumatoid arthritis is an autoimmune disease. It affects many joints, including proximal interphalangeal, metacarpophalangeal, tarsal and temporomandibular joints. It occurs in younger individuals, and females are more often affected than males. Osteoporosis of adjacent bones, subluxation and ankylosis may be found.

Degenerative disease of the vertebrae may involve the apophysial (facets of the vertebrae) or intervertebral joints and is characterized by osteophytes. This condition is so common that it is unlikely to help in a positive identification. Its presence may, however, confirm that the remains are of an elderly individual.

Circulatory Diseases

There are very few circulatory diseases which involve bone. In Perthes' disease, for example, the epiphyses of the head of the femur undergoes aseptic avascular necrosis. It mostly occurs in boys between 5 and 10 years. The head of the femur usually flattens, and the neck thickens. After revascularization the head of the femur is mushroom-shaped with an overhanging margin.

Aneurisms (saccular dilatation of an artery) may erode closely situated bone. Aortic aneurisms, for example, may erode the posterior surface of the sternum. This mostly happens in syphilitic aortitis, but is uncommon.

Metabolic Diseases

Examples of metabolic diseases that may influence bone include endocrine diseases, osteoporosis, Paget's

disease, deficiencies (e.g. vitamin C) and gout. Although nutritional diseases are relatively rare in developed countries, many people, mostly from Africa and Asia, suffer from primary malnutrition. Therefore, diseases such as scurvy (vitamin C deficiency) and rickets (vitamin D deficiency) may be found in skeletal remains of children especially. Skeletal signs of scurvy may show as transverse fractures in the metaphysis near the epiphyses, thin bone cortex and reactive periosteal bone deposition. Rickets leads to bone deformation, softened cranial bones (craniotabes), frontal bossing and the formation of a square head, and a pigeon chest (sternum projecting anteriorly).

Endocrine diseases include, among others, gigantism or acromegaly and dwarfism. Excessive production of growth hormone (somatotropin) leads to gigantism when it starts in the growing years and acromegaly when it begins in adults. Acromegaly is usually due to a tumor of the pituitary and is, therefore, usually associated with an enlarged sella turcica. Since epiphyseal plates are already closed, only the periosteum and cartilage can respond to the stimulation of the growth hormone. The result is elongation of the mandible and bony buildup at the chin, enlarged hands and feet, accentuated supraorbital ridges, enlargement of the facial bones and thickening of the cranial vault. A deficiency of growth hormone may lead to dwarfism.

Osteoporosis is an extremely common disease, in which the bones lose their density and become fragile. It is more common in women, especially after the menopause. This disease is too common and nonspecific to be of any use in personal identification, but its presence usually indicates an older individual.

Paget's disease usually starts after 40 years of age, and becomes more and more prevalent thereafter. It is common in Caucasoids of Europe, North America, Australia and New Zealand, and reaches a frequency of 10% in males and 15% in females by about 90 years of age. It is rare in native Asians and Africans. The disease occurs in three phases: first an initial osteolytic stage; then a mixed osteolytic–osteoblastic stage; and finally a burnt-out osteosclerotic state. Virtually any bone of the skeleton can be affected, and in skeletal remains it is often seen as sclerotic, thickened and enlarged bones. Medullary spaces are diminished, and the distinction between cortex and medulla is obscured. Thickening of the bones of the skull vault is frequently seen. Although signs of Paget's disease on bone are quite obvious, the disease may be asymptomatic or not diagnosed. Symptoms may include pain due to small fractures, increase in head size, and nerve impingement due to foraminal narrowing. Sarcomas may develop on affected bone.

Proliferative Diseases

Proliferative diseases affecting bone include a whole range of conditions, often malignant in nature. Benign tumors of bone include osteomata and osteoblastoma. 'Ivory' or 'button' osteomata are commonly found on bones of the cranial vault (**Fig. 15**), but usually it does not have any clinical significance.

Malignant tumors may be either primary or metastatic. Primary bone malignancies usually occur in younger, and metastases in older individuals. Primary malignant bone tumors include osteosarcoma, chondrosarcoma, and Ewing's sarcoma. Osteosarcoma most frequently occurs in adolescence, and affects males more commonly than females. It is usually associated with enchondral growth, and commonly appears in the metaphyses of the distal femur, proximal tibia and proximal humerus. The appearance may range from total lytic lesions (destruction of bone), to the formation of massive sclerotic bone.

Multiple myeloma is a plasma cell tumor, which manifests as multifocal destructive bone lesions throughout the skeleton. The lesions appear as punched-out, circular defects. This disease occurs in older individuals, and males and females are equally affected.

Various carcinomas metastasize to the skeleton, but they occur less frequently in sarcomas. The most common tumors to metastasize to the skeleton are breast cancer in women and prostate cancer in men. The bones most commonly involved are the spine, femur, ribs, sternum and skull. Fast-growing tumors are usually lytic in nature, and slow-growing tumors osteosclerotic.

Nonspecific Markers of Disease

There are various signs of disease that may act as nonspecific indicators of stress and/or malnutrition.

Figure 15 Ivory 'button' osteomata on the skull (Pretoria skeletal collection). Photo: P. Birkholz.

Although they are not useful in making personal identifications, they give some indication of the living conditions and socioeconomic background of an individual.

Enamel hypoplasia of the teeth present as horizontal lines or pits in the enamel. They are indicative of periods of malnutrition and acute disease/fever during the developmental phase of the teeth which are severe enough to inhibit normal enamel formation. The age at which the defect formed can be calculated from the distance between the hypoplastic line and the cement–enamel junction.

The etiology of transverse radio-opaque lines (Harris lines) of long bones is the same as for enamel hypoplasia. These lines, which can be seen on radiographs, are formed when the normal growth at the epiphyses had stopped, and then started again. The number of lines may serve as an indicator of the amount of stress/disease during the growing years.

Cribra orbitalia refers to pitting in the roof of the orbit, but if it is severe it may also involve the rest of the skull (porotic hyperostosis). Currently the general conclusion seems to be that this is the result of long-standing anemia. The anemia causes bone marrow hyperplasia, resulting in the pitted appearance of the overlying bone.

Conclusions

The importance of health assessment and reconstruction of life from the skeleton has been discussed in a number of forensic anthropology books. It is important to note that assessment of health from skeletal remains is significant in a number of forensic matters; for example, it may give an indication of socioeconomic status. Diseases like osteomyelitis are rare in affluent countries and may more often be associated with individuals of lower socioeconomic status or those living in undeveloped countries. This same assumption may also be made for individuals with inadequate dental care. A correctly diagnosed skeletal lesion of antemortem origin can also lead to the identification of missing people. This is especially true for individuals with previous fractures who are readily recognized by members of their community, or who have detailed medical records. Correct assessment of a lesion can also aid in determining the manner of death. If a claim is made that the deceased died of cranial fractures, it is possible to observe if the fractured area shows evidence of bone remodeling. The presence and amount of growth or healing may lead the investigator to conclude that the fractures have occurred sometime before death.

When advanced disease is present on bones of forensic origin, there may be a question of whether the cause of death was, in fact, natural. Cause of death can be very difficult to determine when these kinds of chronic disease are present. It should be kept in mind that the observed disease may not have been fatal. Since bodies are usually not discovered in unusual places or buried in shallow graves, it is recommended that they are treated as potential unnatural deaths. This is obviously the case in affluent countries, although the situation in undeveloped countries with large rural areas may be different. In these countries it is possible that the recovered remains were those of homeless people dying of natural causes in less inhabited areas.

Antemortem diseases of bone are often difficult to diagnose specifically, and it is recommended that specialist opinions be sought. Radiographs should be taken in all cases to confirm the diagnosis. Modern investigative techniques, such as tomography and magnetic resonance imaging may also help in diagnosis. With the increased development of techniques for DNA extraction from bone, the diagnosis of infectious diseases from the DNA of microorganisms may also be possible.

The most important principle in the evaluation of antemortem disease remains the careful, systematic observation of every bone. It is unfortunate that record keeping of patients by physicians and dentists is not uniformly carried out in every country. Therefore, sometimes even a correct diagnosis of antemortem pathology of a deceased individual may not lead to a positive identification.

See also: **Anthropology**: Overview; Archaeology. **Causes of Death**: Scene of Death; Postmortem Changes; Sudden Natural Death; Blunt Injury; Sharp Injury; Burns and Scalds; Traffic Deaths; Systemic Response to Trauma. **Identification/Individualization**: Overview and Meaning of ID. **Pathology**: Overview. **Anthropology**: Skeletal Analysis; Morphological Age Estimation; Sex Determination; Determination of Racial Affinity; Excavation and Retrieval of Forensic Remains; Bone Pathology and Antemortem Trauma in Forensic Cases; Skeletal Trauma; Animal Effects on Human Remains; Assessment of Occupational Stress; Stature Estimation from the Skeleton.

Further Reading

Apley AG and Solomon L (1993) *Apley's System of Orthopaedics and Fractures*. Oxford: Butterworth Heinemann.

Aufderheide AC and Rodriguez-Martin C (1998) *The Cambridge Encyclopedia of Human Paleopathology*. Cambridge: Cambridge University Press.

Brothwell DR (1981) *Digging up Bones: The Excavation, Treatment and Study of Human Skeletal Remains*. Ithaca, NY: Cornell University Press.

Cockburn TA (1973) *Infectious Diseases: Their Evolution and Eradication*. Springfield, IL: Charles C Thomas.
Cotran RS, Kumar V and Robbins SL (1989) *Robbins Pathologic Basis of Disease*. Philadelphia: WB Saunders.
İşcan MY and Kennedy KAR (eds) (1989) *Reconstruction of Life from the Skeleton*. New York: Alan R. Liss.
Krogman WM and İşcan MY (1986) *The Human Skeleton in Forensic Medicine*. Springfield, IL: Charles C Thomas.
Ortner DJ and Putschar WGJ (1981) *Identification of Pathological Conditions in Human Skeletal Remains*. Washington, DC: Smithsonian Institution Press.
Praemer M, Furner S and Rice DP (1992) *Musculoskeletal Conditions in the United States*. Park Ridge, IL: American Academy of Orthopedic Surgeons.
Resnick D and Niwayama G (1988) *Diagnosis of Bone and Joint Disorders*. Philadelphia: WB Saunders.
Rothschild BM and Martin LD (1993) *Paleopathology: Disease in the Fossil Record*. Boca Raton, FL: CRC Press.
Steinbock ST (1976) *Paleopathological Diagnosis and Interpretation: Bone Diseases in Ancient Human Populations*. Springfield, IL: Charles C Thomas.
Zimmerman MR and Kelley MA (1982) *Atlas of Human Paleopathology*. New York: Praeger.

Determination of Racial Affinity

M Y İşcan, Adli Tıp Enstitusu, Istanbul Univesitesi, Istanbul, Turkey
S R Loth and **M Steyn**, Department of Anatomy, University of Pretoria, Pretoria, South Africa

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0757

Introduction

An important aspect of forensic anthropology is the determination of the racial affinity of an unknown skeleton. Today, the world has become so accessible that there are few places where the populace is totally homogeneous. Even in countries where there is minimal migration from other geographic regions, it is not uncommon to see people of different origins passing through as tourists or in exchange programs. Therefore, the geographic isolation of the past is rare. Currently, the assessment of race is complicated by every kind of admixture as well as the variability that exists within races at the population level.

There have been numerous attempts to explain the origin of populations residing in different geographic regions. There are also textbooks that focus on what races are, how they are formed, and how they change. Even the use of the term 'race' has been a sensitive issue among many scientists because of the unwarranted social connotations that have been associated with it. Therefore, when dealing with the complex and often controversial subject of race, it is essential for experts to convey exactly what is meant. Recent efforts to deal with the controversy and negative image of this designation have led to the use of incorrect and misleading terminology. Some anthropologists have begun misusing the term 'ethnicity' (which refers to cultural issues like nationality, religion, etc.) when discussing genetically based biological characteristics readily observable in the flesh and in the skeleton. A more appropriate term is 'racial phenotype' because it accurately describes what forensic anthropologists determine when they examine skeletal remains.

Another problematic area is the terminology used to categorize racial phenotypes. To meet the needs of forensic case work, and avoid confusion, we use the three traditional classifications: Caucasoid (white), Mongoloid (east Asian, Amerindian), Negroid (black). Today, a number of anthropologists substitute the term 'Asian' for Mongoloid, but this is inaccurate because not all Asians are Mongoloid (e.g., Asian Indians, Turks, Iranians) and not all Mongoloids are Asian (Amerindians). Most humans can be roughly classified as a member of a population within these designations, or some admixture thereof. Unfortunately, this is not always a simple matter, especially where complex admixtures are concerned. In addition, skeletal features give no clue to soft tissue characteristics such as the precise shade of skin color or eye shape.

A single individual can represent the three major racial phenotypes as well as any number of populations within them. The golfer Tiger Woods is a well-known example. At the population level, a good model of a highly mixed gene pool is found in the eastern nations of South America, such as Brazil, where there is much African, European and indigenous Mongoloid (Amerindian) admixture. Western countries like Peru and Chile have a long history of white and aboriginal intermixture, whereas others, like Argentina and Uruguay, have remained primarily white. The United States and South Africa are also centers of mixed populations from all around the world. With this kind of admixture, genes for original ancestral traits may be transmitted unpredictably and randomly to the next generation. However, there do seem to be some characteristics that remain 'typical' of obvious ancestral traits, and these are valuable markers that forensic anthropologists use in their investigations to establish the identity of an individual.

People are only recognizable to others by their outward appearance, and this is often dominated by the characteristics of a particular racial phenotype or sometimes an obvious combination thereof. The medicolegal expert must therefore begin investigations with these population specific facial traits even if they do not reflect someone's actual genetic constitution. Thus, the purpose of this article is to investigate the ways in which a given individual or victim of a crime can be classified as Caucasoid, Mongoloid or Negroid from skeletal remains.

Methodologies

As in the determination of sex, race assessment from the skeleton can be carried out using both morphological and metric traits. Morphological traits involve shape and size differences that can be assessed with the naked eye. This type of analysis requires consistency in observation and improves with practice. Therefore, the forensic anthropologist who relies on the visible configuration of a particular structure to establish race must have an advanced education and considerable experience in observing and understanding skeletal variation between individuals, populations, and age groups. Metric determination of race is based on the use of selected measurements that show statistically significant population differences. Metrics require both statistical expertise and a thorough knowledge of osteometric techniques and principles. The most commonly used statistical approach for separating two or more groups is uni- or multivariate discriminant function analyses.

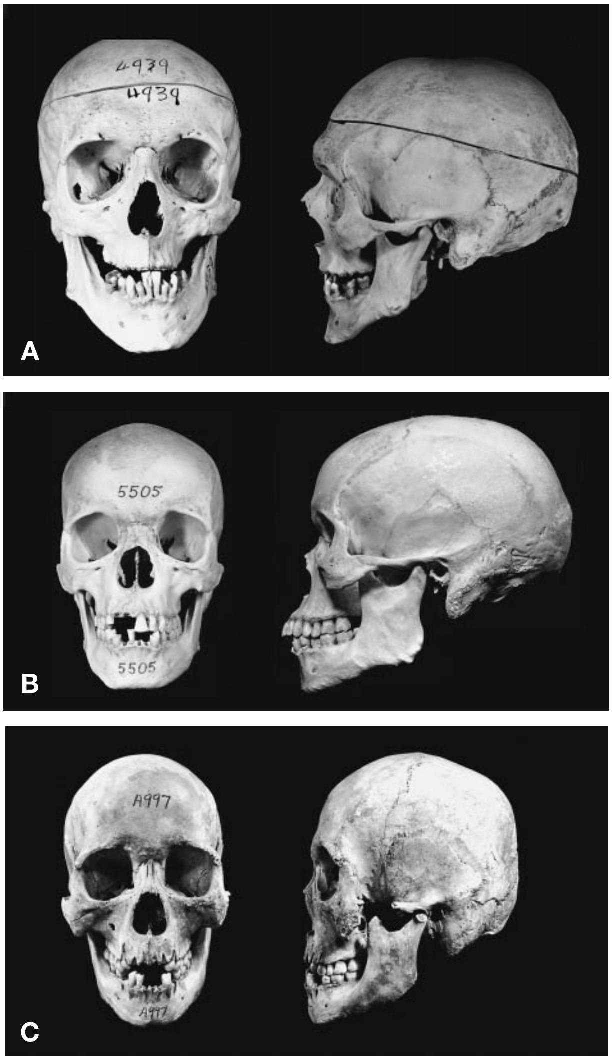


Figure 1 Frontal and profile views of Caucasoid (**A**), Negroid (**B**) and Mongoloid (**C**) skulls.

Morphological assessment

Morphology is shape or configuration. A morphologic trait is one that can be evaluated by observation rather than through measurements. Over the years, race determination has relied primarily on the morphologic features of the skull and facial skeleton since these provide consistently reliable results in the majority of cases.

Some race-specific traits may not be present or fully developed in the young until they reach sexual maturity. The only exceptions to this seem to be the traditionally race-related morphology of the nasal sills, and alveolar prognathism. The wide nasal form and alveolar prognathism in Blacks can be clear even in infants. However, one must keep in mind that the nasal bridge is not well developed in any infant.

Morphologically, race differences are most pronounced and observable in the skull, with the facial skeleton being the most diagnostic portion (**Fig. 1**). **Table 1** lists some craniofacial features that have been linked to race in the three major groups. Following the process of dental analysis, a craniofacial race complex can be identified to separate Whites from Blacks. In Whites, this complex includes a high, wide, skull, rounded sagittal contour, low cheek, short, gracile zygomatic arches, orthognathic face, narrow interorbital distance, high nasal bridge, sharp nasal sills, and narrow nasal aperture. The Black complex is characterized by a long, low, narrow skull, flat sagittal contour, long, robust zygomatic arches that project laterally (relative to the narrowness of the head), alveolar prognathism, wide interorbital distance, receding chin, low nasal bridge, smooth, guttered nasal sills, and wide nasal aperture. The Mongoloid complex features a rounder skull, anterior and laterally projecting zygomatic bones, flat face with little projection, shallow nasal root, and shovel-shaped incisors. The most distinctive Mongoloid feature is

Table 1 Craniofacial traits of the three major human races

Characteristics	*Caucasoid*	*Negroid*	*Mongoloid*
General appearance	High, wide	Long, low	Rounded
Skull length	Varied	Long	Long
Skull breadth	Varied (often broad)	Narrow	Broad
Skull height	Medium to high	Low	Medium
Sagittal contour	Round to arched	Flat	Arched
Face breadth	Narrow to wide	Narrow	Very wide
Face height	Medium to high	Low	High
Cheek	No projection	Some lateral projection	Anteriorly projected
Orbital opening	Round to angular	Rectangular	Rounded
Facial profile	Orthognathic	Prognathic	Medium to flat
Nasal opening	Narrow to medium	Wide	Narrow to wide
Lower nasal margin	Sharp	Smooth and/or guttered	Sharp
Nasal profile	Straight	Downward slant	Straight
Palate shape	Narrow to wide	Wide	Moderately wide

Modified from Krogman and İşcan (1986).

the cheekbones. They exhibit high malar projection, both anteriorly and laterally, as well as a malar tubercle at the inferior aspect of the zygomaxillary suture. In some cases the bizygomatic breadth (measurement of facial width) may be as wide as the breadth of the skull. Also, shovel-shaped incisors are much more frequent in Mongoloids, as are taurodont molars. Interestingly, shovel-shaped incisors are also common in Caucasoid residents of the Asian Indian subcontinent (e.g. Indians, Pakistanis).

One of the best racial indicators is alveolar prognathism. In Blacks the tooth bearing portions of the jaws protrude noticeably in contrast to the straight or orthognathic face of whites and Mongoloids. Nasal configuration is also very telling. A short, wide, low bridged, flat nose with smooth and often guttered sills is an extremely reliable marker of Negroid affinity. A long, narrow, protruding nose with sharp sills points quite reliably to a white person.

Race differences are also apparent in the mandible and dentition. The tall, thin, gracile ramus noticeably narrows in the middle in Whites as opposed to the robust, rectangular or almost squared ramus in Blacks. This contrast is distinctly visible in the lateral photographs in Fig. 1. A wide, vertical ramus is associated with Mongoloids. Blacks (especially Africans) are also more likely to have fully erupted third molars, supernumerary teeth (e.g. fourth molars), and receding chins than other groups. Moreover, Blacks retain their dentition with age much better than do Whites and are less likely to have dental work (fillings, caps etc). Finally, there are significant differences in tooth size, with Blacks having the largest, and Whites the smallest dentition. Australian aborigines have the largest teeth of any living humans. However, these are only visually apparent to the most experienced observers.

Although there may be considerable variability, some generalizations can be made about head shapes. Whites tend to have a high, wide head, Mongoloids are more rounded, and blacks often exhibit long, low crania. Other morphological features have also been described. The nasal bridge has been observed to have a distinct distribution among the races. The Caucasoids have a 'church with steeple' shape (high and pinched), Negroids a 'Quonset hut' (low and rounded), and Mongoloids a 'tented' appearance (low to moderate with relatively straight sides).

Morphologic differences are less obvious in the postcranial skeleton. One exception is femoral curvature which differs noticeably in that American Blacks have a much straighter femur in contrast to the pronounced anterior curvature in Whites and Mongoloids. The pelvis is also quite disparate between Whites and Blacks. The white pelvis has been described as a broad basin with a lower symphysis than Blacks to compliment a broad torso, and in Blacks as the pedestal of a narrow torso. There are significant shape differences in the pelvic brim that give Whites a wider, rounder (brachypellic) inlet and a narrower (dolichopellic) inlet in blacks. However, this is not always easy to distinguish visually and the metric approach to race assessment in the postcranial skeleton is preferred.

Determination of race in children has focused on the variation of several dental characteristics. Using 10 deciduous dental characteristics, it is sometimes possible to determine racial affinity in children. The Mongoloid dental complex includes six traits with higher than average frequencies. One of these is the

shovel-shaped morphology in the upper central and lateral incisors as well as canines (**Table 2**). The percentage of shovel-shaped incisors in i1 is about 77 in the Japanese, 10 in American Blacks and zero in Whites. The same trend is also seen in i2. In the Caucasoid complex, Carabelli's cusp is as high as 35% in Whites and about 12% in Blacks. The canine breadth index (upper canine crown mesiodistal diameter × 100/upper central incisor mesiodistal diameter) is also higher in Caucasoids and Negroids than Mongoloids. Some trace this to the fact that American Blacks and Whites have had more opportunity for admixture during the last 250 years or so.

When clear, morphologic traits are very effective for racial assignment Interestingly, many of these traits have also pointed scientists in the direction of quantification as another approach to the definition of certain formations. In this way, some morphologic traits have formed the basis of many metric techniques. For example, projection of the alveolar region of the face can be observed as either orthognathic, mesognathic or prognathic by looking at how far this region projects forward, but it can also be assessed by measuring the angle of the face.

Metric differences

The primary focus of metric assessment is size and proportion. Because there are significant size differences between races and the populations within them, metric techniques can be used to separate them. In cases where morphologic variants are not clear, or the facial skeleton is not available, measurement-based methods can provide good levels of discrimination. This technique necessitates a thorough knowledge of skeletal landmarks, proper equipment, and precise measuring skills. Furthermore, there can be complex combinations of shape and size differences between populations that are not obvious to the eye, and these may be quantified and evaluated by using a set of measurements. The statistical technique most commonly used is discriminant function analysis. This approach assumes that human variation spans a continuum across space and populations, but concentrations of people with similar features can be found towards the centers, whereas at the peripheries, there may be an overlap with neighboring groups. Because of the overlap, it is rare to be able to divide humans into distinct groups with 100% accuracy. Optimal diagnosis is obtained in areas where there are two major races, with little admixture and no intervening populations. Discriminant function analysis has become very popular, especially during the last 40 years and has been applied to many parts of the skeleton for purposes of determining both race and sex.

In order to test variation among three biologically diverse samples, data were obtained from American Whites and Blacks in the Terry collection (United States) and Japanese (Mongoloids) from the Jikei collection (Japan). They were analyzed to determine the extent to which these three groups can be differentiated from each other (**Table 3**). The first set of discriminant function tests assesses the differences between two samples at a time, e.g. American Whites and Blacks (**Table 4**). The second set is a multivariate discriminant function statistic that determines if an unknown skull can be assigned to one of the three groups (**Table 5**).

Table 3 lists descriptive statistics for 12 cranial measurements of both sexes of Japanese and American Whites and Blacks using a sample ranging in size from 29 to 44 specimens per race–sex group. Blacks have the greatest cranial length, Japanese, the largest

Table 2 Population variation in deciduous dental crown characteristics

	Frequency (%)		
Crown characteristics	*Whites*	*Blacks*	*Japanese*
Mongoloid complex			
Shovel shape (upper i1)	0	10	76.6
Shovel shape (upper i2)	0	15	93.3
Deflecting wrinkle (lower m2)	13	19.1	55.6
Protostylid (lower m2)	14.5	17	44.7
Seventh cusp (upper m2)	41.8	46.8	73.1
Metaconule (upper m2)	3.5	9.5	41.8
Caucasoid complex			
Carabelli's cusp (upper m2)	35.1	11.8	11.9
Canine breadth index (upper c)	106.3	107.8	101.5

Modified from Hanihara (1967).

Table 3 Descriptive statistics, univariate F ratio, and significance of differences among American Blacks and Whites, South African Blacks and Whites, and Chinese and Japanese

	Whites		*Blacks*		Japanese		*F ratio and statistical significance between*		
Dimensions	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*White–Black*	*White–Japanese*	*Black–Japanese*
Male	35.0		32.0		44.0				
Cranial length	182.2	7.47	186.9	5.65	178.5	7.74	8.45[b]	4.53[a]	27.23[c]
Cranial breadth	142.9	5.06	138.3	4.43	142.2	5.87	15.54[c]	0.39	9.55[c]
Max. Frontal b.	123.7	5.81	119.4	4.48	119.2	5.39	11.10[c]	12.67[c]	0.04
Min. Frontal b.	95.9	5.50	97.5	4.45	95.4	5.18	1.79	0.16	3.49
Bizygomatic b.	130.7	4.62	131.9	7.30	135.6	5.70	0.65	16.46[c]	5.93[a]
Basion-nasion	100.4	4.17	101.4	4.14	101.4	3.92	0.98	1.21	0.00
Basion-bregma	133.8	8.15	131.0	6.88	138.4	5.06	2.29	9.71[c]	29.76[c]
Basion-prosthion	93.3	5.16	103.2	8.81	94.7	4.98	32.36[c]	1.65	28.25[c]
Mastoid height	30.3	3.04	32.6	3.58	32.9	2.33	8.55[c]	19.61[c]	0.20
Biasterionic b	111.8	5.77	109.0	5.85	107.9	3.84	3.97[a]	13.16[c]	0.99
Nasal height	51.7	2.63	53.1	4.92	52.0	2.49	2.01	0.13	1.75
Nasal breadth	23.9	1.59	27.2	2.19	26.2	1.90	47.93[c]	30.49[c]	4.48[b]
Female	31.0		30.0		29.0				
Cranial length	177.2	8.09	178.0	6.80	173.2	7.33	0.21	3.85[a]	6.79[b]
Cranial breadth	139.2	5.50	133.6	5.82	137.3	6.49	14.90[c]	1.58	5.17[a]
Max. Frontal b.	118.5	5.71	113.0	4.68	113.1	6.06	17.16[c]	12.52[c]	0.01
Min. Frontal b.	93.4	4.47	93.8	3.20	91.2	4.70	0.14	3.29	5.86[a]
Bizygomatic b.	122.1	3.14	124.4	5.48	128.9	5.40	4.33[a]	36.86[c]	10.09[c]
Basion-nasion	96.0	4.75	96.1	4.08	97.7	4.27	0.00	2.09	2.23
Basion-bregma	129.2	4.74	124.1	6.59	132.4	5.86	12.23[c]	5.29[a]	25.96[c]
Basion-prosthion	87.4	5.18	98.4	6.91	94.1	5.16	49.55[c]	25.05[c]	7.28[b]
Mastoid height	28.0	2.70	28.4	3.10	29.0	2.98	0.24	1.87	0.64
Biasterionic b	109.6	6.76	104.4	5.14	104.9	4.90	11.44[c]	9.47[c]	0.14
Nasal height	49.8	3.41	49.2	3.10	50.1	3.06	0.53	0.08	1.08
Nasal breadth	23.2	1.76	26.1	2.41	26.2	2.16	28.35[c]	33.75[c]	0.01

[a] Significant at $P < 0.05$ level.
[b] Significant at $P < 0.01$ level.
[c] Significant at $P < 0.001$ level.

Table 4 Population determination from paired combinations of Japanese and American White and Black crania and their classification accuracies for males and females

Dimensions	*White–Black coefficient*	*White–Japanese coefficient*	*Black–Japanese coefficient*
Males			
Cranial length		−0.0403828	−0.0876780
Cranial breadth	−0.0723247		0.1273584
Max. frontal breadth		−0.0955337	−0.1089742
Bizygomatic breadth		0.1231765	
Basion–nasion length	−0.1097166		
Basion–bregma height		0.0575340	0.1139350
Basion–prosthion length	0.1125220		
Mastoid height		0.1511110	
Biasteronic breadth	−0.0623639	−0.1207576	
Nasal breadth	0.3694600	0.2081077	
Constant	7.6988166	−2.2330051	−4.3549907
Sectioning point	−0.5850100	−0.9330200	−0.7740400
% Accuracy	90.4100000	88.3700000	86.0500000
Females			
Cranial length			−0.0881222
Cranial breadth	−0.1237725		
Max. frontal breadth		−0.0702601	
Min. frontal breadth			−0.1585874
Bizygomatic breadth		0.1706030	0.1416418
Basion–nasion length	−0.1180378		
Basion–bregma height			0.1269521
Basion–prosthion length	0.1514658	0.0686073	
Mastoid height	0.1536866		
Biasteronic breadth		−0.0883313	
Nasal breadth	0.2334410	0.1919761	
Constant	4.0805893	−14.7170241	−4.0549844
Sectioning point	−0.7309400	−0.6727900	−0.5958100
% Accuracy	98.48	95.08	89.19

Table 5 Population determination (three group) from Japanese and American White and Black crania and their classification accuracies for males and females

	Males			*Females*		
Dimensions	*Function 1 coefficient*	*Function 2 coefficient*		*Function 1 coefficient*	*Function 2 coefficient*	
Cranial length	−0.042814	0.046077		−0.022793	−0.074372	
Cranial breadth	0.085895	−0.102257				
Max. frontal breadth	−0.159036	0.076478		−0.089974	−0.034754	
Bizygomatic breadth	0.087832	−0.035634		0.154269	0.076963	
Basion–bregma height	0.049601	−0.065688		0.039990	0.125313	
Basion–prosthion length	0.040147	0.079534		0.098570	−0.075530	
Biasteronic breadth	−0.105740	−0.019376		−0.088455	0.049045	
Nasal breadth	0.253434	0.245891				
(Constant)	−2.358715	−1.475785		−9.859351	−6.808660	
Group centroids	Whites	Blacks	Japanese	Whites	Blacks	Japanese
Function 1	−1.438670	0.117840	1.058690	−1.641180	0.542170	1.193500
Function 2	−0.548220	1.591600	−0.721440	0.257780	−1.159270	0.923690
Classification						
Accurate N of total N	34/35	24/32	34/44	28/31	27/31	23/29
% accuracy	97.1	75.0	77.3	90.3	87.1	79.3

bizygomatic breadth, and Whites, the largest maximum frontal and biasterionic breadths. Nasal width is the narrowest in Whites and widest in Blacks. Analysis of variance (F ratio) showed statistically significant differences between all groups. Using stepwise discriminant function statistics, only four to seven male and four to five female dimensions were selected to provide the maximum separation among the three samples (Table 4). It must be kept in mind that this type of assessment of an unknown individual is based on pair group comparisons (e.g. White/Black, White/Japanese, Japanese/Black), rather than as one of all three. Racial affinity can be calculated by multiplying each coefficient by the value of that dimension plus the constant and adding them to obtain the discriminant score. The score is then compared with the sectioning point calculated to separate each pairing of the three groups. The same table also provides the percentage classification accuracy. On average, females were more accurately classified than males. In white–black comparisons, for example, accuracy is about 90% for males and 98% for females.

When three (unpaired) groups are considered, discriminant function statistics can be expected to yield lower separation accuracy because the effectiveness of some variables that are particularly good discriminators between two groups may be diminished by the intervening population. Using the same samples (Japanese, American Whites and Blacks), Table 5 presents discriminant function coefficients, and classification accuracy. Of nine variables that differ significantly, eight were selected by the multivariate stepwise discriminant function (all but mastoid height) for males and six (all but mastoid height, cranial breadth, and nasal breadth) for females. Accuracy averaged 82.8 for males and 85.7 for females. As in the univariate function (Table 3), both sexes of Whites were more successfully diagnosed than the others. Furthermore, there was no particular tendency in the distribution of misclassified cases. Although nasal breadth was the most distinguishing characteristic between Blacks and Whites, the addition of the Japanese sample (with an intermediate nose width closer to that of blacks) canceled the effectiveness of that variable in females. As expected cranial length showed more population variation than cranial breadth in this three-group analysis. **Figure 2** shows the territorial placement of three samples for both sexes. Calculation of discriminant scores for both functions 1 and 2 can be made by multiplying each dimension by its coefficient and adding them up along with the constant. The resulting two scores can be plotted in this figure or compared with the group centroids given in Table 5.

There is an important caveat to the use of discriminant formulas on other populations until thorough tests have been conducted. Although the Japanese are, for example, Mongoloid, they may not be representative of other Mongoloid groups such as Amerindians or even the neighboring Chinese. The same applies to using American Black standards for the much smaller southern African Blacks since there are significant mensural differences between them.

Similar studies were also done on the postcranial skeleton such as the pelvis and long bones of American Whites and Blacks. It was shown that when several standard measurements from the pelvis, femur and tibia were combined, a person could be classified as White or Black with about 95% accuracy. The pelvis alone can differentiate Whites from Blacks by as much as 92%. The femur and tibia are less discriminating but, individually and in combination, can still produce up to 88% separation between American Whites and Blacks.

There are well-known proportional differences

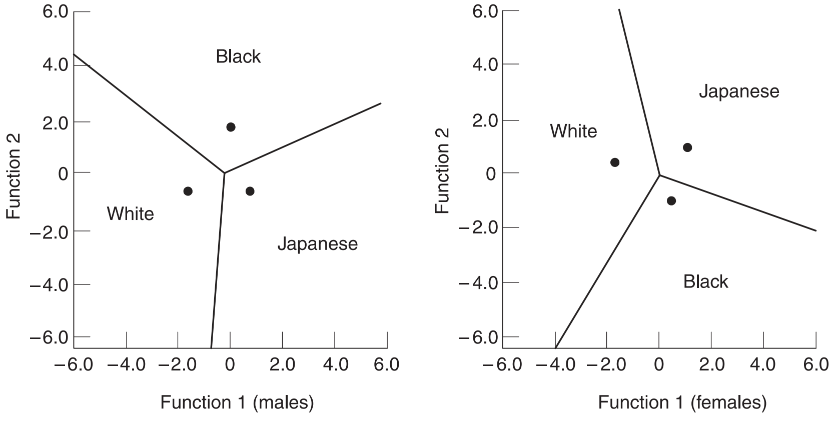


Figure 2 Population placement (via 3 group analysis) of American White and Black, and Japanese of both sexes using cranial dimensions. (* indicates group centroids; see Table 5.)

among the races based on limb to height ratios. In Blacks, limbs are longest relative to height, and the reverse is true for Mongoloids. Whites are intermediate. This variation, which has been associated at least in part with climate, is reflected in the fact that it was necessary to calculate race-specific regression formulas to estimate stature from the long bones. This proportional variation also explains why the sternal end of the rib demonstrates White/Black racial dimorphism with Whites being significantly larger. This is consistent with other torso components such as those of the pelvis where Whites are also larger in all dimensions tested.

Recent research on the metric analysis of race in subadults has revealed that the dimensions of the calvarium, especially cranial breadth, resembles the adult pattern of differences between Whites and Blacks after the age of 5 years. Ideally, these results will be confirmed by testing on other, larger samples.

Discussion

The determination of race is a complicated task in part because humans are biologically dynamic, through both adaptation and interaction. Aspects of culture, such as advancing technology and changing attitudes, are also major factors for they bring together genetically and geographically diverse peoples. In many cases culturally and socially assigned race may not reflect the true genetic make-up of the deceased. For example, no matter how much white admixture there is in the offspring of a White–Black union, dark skin color usually dominates the physical appearance, and the general public considers this to be the race of that child. The difficulty increases as cultures change and transportation improves. Although not everyone has access to airlines, few people are more than 24 hours apart. There are, however, certain groups who have maintained their isolation from others much more successfully and their gene pool is less mixed than those of other contemporary populations. Some of these can be found in the Kalahari desert of southern Africa, remote regions of China and Siberia, islands of the Philippines, Indonesia, and New Guinea, northwestern populations of Asia, and even bands of Inuits.

As in all skeletal assessments, both morphological and metric examinations must be carried out meticulously and the results carefully considered. There are always individual differences in the size or shape of a particular structure and equally important is the interobserver variation that forms the basis for interpreting what is seen. It must also be kept in mind that expertise in one population does not always make a person equally expert in analyzing variation in others. Discriminant function statistics have also been problematic in terms of applicability to populations of similar racial origin. It is highly unlikely, for example, that the standards based on American Blacks will be applicable to their African counterparts.

In the medicolegal arena, racial placement is an essential piece of the identification puzzle. Even the most skilled expert could not attempt facial reconstruction without this knowledge as a necessary guide to soft tissue morphology. Moreover, race differences can interfere with the accurate assessment of other demographic characteristics like age, sex and stature. For example, race affects both age and sex determination in the rib. Ribs of Blacks are known to age at a different rate and pattern from those of Whites and thus may be assigned to the wrong age phase using White standards. And, if sexing is attempted using metric values for White ribs, the vast majority of Black males would be classified as female. Similar results would be expected using metric sexing standards for American Whites on Asian Indians or figures for American Blacks on southern African Blacks. The proper regression formula must be used for statural estimation from long bones because of variation in body proportions. Height will be underestimated in a White if the formula for Black is used and overestimated in a Black using the formula for Whites.

As noted above, the best one can do in determining the racial phenotype of a contemporary person is to assign the individual as Caucasoid, Negroid, Mongoloid or some admixture thereof. Although diagnostic results are often accurate for most adults, further research is needed to improve the identification of race in infants and children.

In conclusion, it is nearly impossible to establish the identity of skeletal remains without determining race. Forensic examiners must be extremely cautious in making a decision on this aspect of the identification of unknown remains. Incorrect racial assignment will thwart identification by eliminating from consideration the population containing the individual in question. Forensic anthropologists and law enforcement agents must be careful not to rule out missing individuals who do not show ‘typical’ features for a given racial phenotype.

See also: **Anthropology**: Overview; Archaeology; Skeletal Analysis; Morphological Age Estimation; Sex Determination; Excavation and Retrieval of Forensic Remains; Bone Pathology and Antemortem Trauma in Forensic Cases; Skeletal Trauma; Assessment of Occupational Stress. **Identification/Individualization**: Overview and Meaning of ID.

Further Reading

Brace C (1995) Region does not mean 'race' – reality versus convention in forensic anthropology. *Journal of Forensic Sciences* **40**:171–175.
Brues AM (1992) Forensic diagnosis of race: general race vs. specific populations. *Social Science and Medicine* **34**:125–128.
Brues AM (1990) *People and Races*. Prospect Heights, IL: Waveland Press.
Coon CS and Hunt EE (1965) *The Living Races of Man*. New York: Alfred A. Knopf.
El-Najjar MY and McWilliams RK (1978) *Forensic Anthropology: The Structure, Morphology, and Variation of Human Bone and Dentition*. Springfield, IL: Charles C Thomas.
Gill GW and Rhine S (eds) (1990) *Skeletal Attribution of Race: Methods for Forensic Anthropology*. Anthropological Papers no. 4. Albuquerque: Maxwell Museum of Anthropology.
Hanihara K (1967) Racial characteristics in the dentition. *Journal of Dental Research* **46**:923–928.
Howells WW (1970) Multivariate analysis for the identification of race from crania. In: Stewart TD (ed.) *Personal Identification in Mass Disasters*, pp. 111–122. Washington, DC: National Museum of Natural History.
İşcan MY (1990) Race determination from the postcranial skeleton. *Adli Tip Dergisi (Journal of Forensic Medicine)* **6**:129–140.
İşcan MY and Steyn M (1998) Craniometric assessment of population affinity in South African blacks and whites. *International Journal of Legal Medicine*. (in press).
Kennedy KAR (1995) But professor, why teach race identification if races don't exist? *Journal of Forensic Sciences* **40**:797–800.
Krogman WM and İşcan MY (1986) *The Human Skeleton in Forensic Medicine*. Springfield: Charles C. Thomas.
Loth SR (1990) A comparative analysis of the ribs of Terry Collection blacks. *Adli Tip Dergisi* **6**:119–127.
Loth SR (1996) *Sexual dimorphism in the human mandible: A developmental and evolutionary perspective*. PhD dissertation. Johannesburg, South Africa: University of Witwatersrand.
Molnar S (1992) *Human Variation: Races, Types, and Ethnic Groups*. Englewood Cliffs, NJ: Prentice Hall.
Sauer NJ (1992) Forensic anthropology and the concept of race: if races don't exist why are forensic anthropologists so good at identifying them. *Social Science and Medicine* **34**:107–111.
St. Hoyme L and İşcan MY (1989) Determination of sex and race: accuracy and assumptions. In: İşcan MY and Kennedy KAR (eds) *Reconstruction of Life from the Skeleton*, pp. 53–94. New York: Wiley-Liss.
Todd TW (1929) Entrenched Negro physical features. *Human Biology* **1**:57–69.

Excavation and Retrieval of Forensic Remains

M Steyn and **W C Nienaber**, Department of Anatomy, University of Pretoria, Pretoria, South Africa
M Y İşcan, Adli Tıp Enstitusu PK10, Cerrahpasa, Istanbul, Turkey

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0758

Introduction

The need for an archaeological assessment of a location where human remains are expected to be found arises frequently. Mass murders in various countries such as Bosnia may also require archaeological expertise. Sometimes a particular grave in a cemetery may have to be excavated, in order to observe whether it is disturbed and if the remains belong to the person who was reportedly buried there. Forensic archaeology entails the application of standard archaeological techniques, modified to meet the needs of crime scene processing where skeletons or buried bodies are present.

In the past it has often happened that burials were excavated by large ground-moving equipment and shovels. Surface remains were collected without paying attention to the distribution of the remains and other taphonomic factors. It is clear, however, that much information is lost in this way. Scientific excavation and retrieval require some basic training, and police departments are slowly incorporating archaeology courses in their training schedules. Retrieval, excavation, analysis and personal identification of decomposed remains and their associated artifacts require a multidisciplinary approach involving additional experts such as pathologists, botanists and entomologists. The purpose of this article is to describe forensic archaeological principles. The discussion covers scientific excavation techniques, discovery, recovery and processing of human remains and artifacts, as well as advising the police about the circumstances in which the remains are buried.

Principles of Archaeology

Two basic principles of archaeology are stratigraphy and superposition. Stratigraphy refers to the formation of strata or layers. It is the sum total of the processes whereby the layers are accumulated. Superposition implies that the oldest evidence is deposited first, and located at the deepest layer. The formation of stratigraphy can be described as the opposite of the

natural process of erosion, but erosion can play a role after the various matrices have been laid down. Water and wind erosion, as well as plant, human and animal activity can alter the deposits. It is impossible to dig a grave, and then cover it up in such a way that the exact, original stratigraphy is retained. Superposition gives an indication of the relative order in which objects were placed in a grave, thus, the objects placed in a grave last are discovered first.

Location of Skeletal Remains

One of the most difficult aspects in forensic archaeology is to find the buried body. Graves are often found by accident or as a result of construction work. An eye witness or informant can also lead the police to a site where there may be human remains (buried or on the surface). Surface indicators are often destroyed in older burials, and it is not always possible to determine whether the grave is of forensic or archaeological nature. Situations may also arise where relatives of a deceased person may claim that the burial was carelessly disturbed and the remains scattered by the cemetery authority during construction or relocation. They may request that the grave is opened to see if it is disturbed and whether all parts of the body are still there.

There are several techniques to locate buried bodies ranging from simple observation to the use of sophisticated equipment. In general, surface changes in soil and vegetation may indicate a grave. These include the following.

1. Soil compaction: although the infill of the grave would probably have been leveled with the surface at the time of burial, the soil will compact after some time to form a concave area, or double depression, known as the primary and secondary depressions. These depressions are not only formed because of the compaction of the infill but also as a result of the decomposition of the buried body. The largest volume to collapse during decomposition is the thorax (secondary depression). This secondary depression occurs in the upper third of the grave's surface, and is usually deeper than the total concave area which denotes the approximate outline of the grave pit (**Fig. 1**). The depth varies according to the type of soil, the depth of the grave, and the amount of water present. This depression is the most obvious in the first months after the burial.
2. Disturbed vegetation: this includes the plants on the surface of the grave itself and surrounding areas where the excavated soil (upcast) was thrown when the remains were interred (Fig. 1). Usually the area directly above the grave may be void of plants for some time, although a shallow grave, with no wrapping around the body, may stimulate growth due to the nutrients originating from the decaying body (**Fig. 2**). The presence of pioneer plants, e.g. weeds, occurring in a localized area may also be indicative of a fresh grave. Weeds are usually the first plants to appear because they are fast growers, and they can often be distinguished from the surrounding vegetation.
3. Disturbed soil: usually the soil from the surface

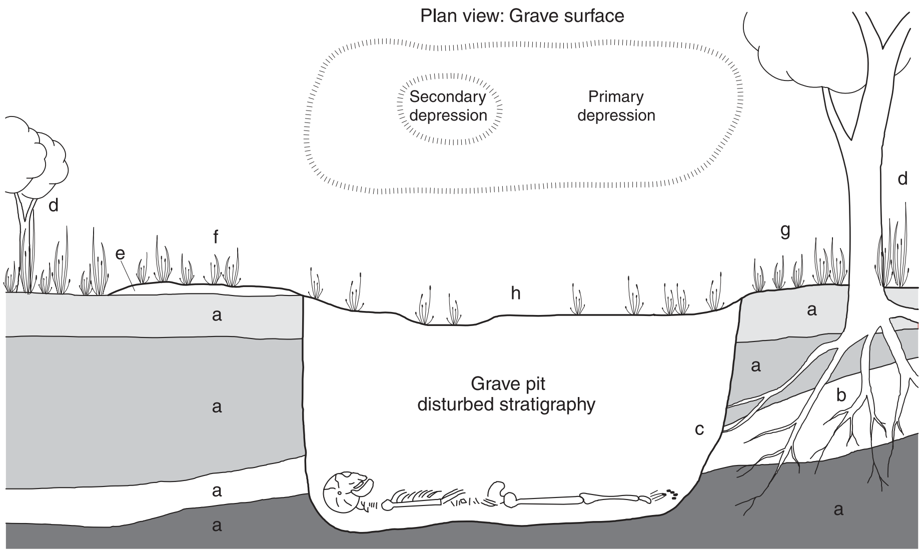


Figure 1 Indications of the presence of a grave. a, Undisturbed stratigraphy; b, undisturbed plant roots; c, disturbed plant roots; d, original undisturbed vegetation; e, upcast remaining on the surface; f, g, different vegetation in the area trampled and disturbed during the original excavation of the grave; h, new plant growth.

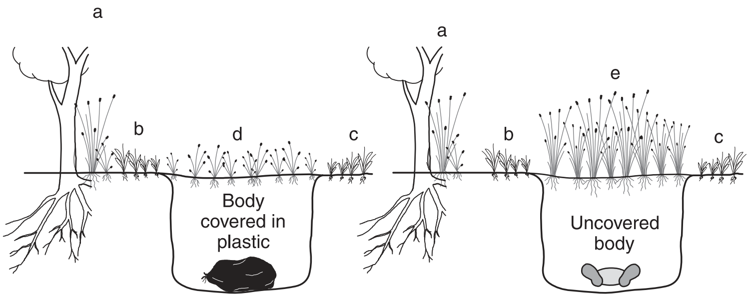


Figure 2 Plant growth in the vicinity of the remains. a, Original vegetation; b, c, different vegetation in the area trampled and disturbed during the original excavation of the grave; d, area of new growth on the grave surface; e, area of lush new growth on the grave surface due to the availability of nutrients from the decomposing body.

and that originating from the grave pit mixes, which leads to obvious differences in color and texture with the surrounding matrix.

These factors are, of course, influenced by the environment. Unless the event occurs soon after harvesting in a farmland these changes may be difficult to observe because of regular plowing. The same is true of sandy soil, where the sides of the grave may collapse and obscure the depression of the grave's surface.

Specially trained cadaver dogs may be used to locate buried remains. They smell the gas formed by the process of decay, and would therefore be most effective shortly after death. A metal detector is very useful to start the surface examination. The ground may also be probed to locate areas of different compaction. Several methods for remote sensing of graves exist, such as ground-penetrating radar, infrared and aerial photography, electromagnetic radiation and the use of microwaves.

Equipment

The basic toolkit of any (forensic) archaeologist includes equipment for setting up the grid, excavation (to locate and unearth the remains) and once found, removal with finer tools. A suggested list is shown in **Table 1**. This list may be expanded, where possible, to include a GPS (global positioning system) instrument to ascertain the locality of the grave and a metal detector. For cases where advanced decomposition has taken place but much soft tissue is present, a body bag may be needed both to reduce the offensive odor and to avoid any infectious disease agent dispersement and to retain/preserve evidence (e.g. bullets).

Recovery of the Remains

All evidence must remain *in situ* until documented. A clear, concise, meaningful, written description of the pertinent aspects of the crime scene is the most important method of documentation. The recovery should be done in such a fashion that the integrity of the evidence is maintained in order for the information to be acceptable in a judicial or criminal investigation procedure. Context is provided via documentation, and contamination must be avoided. All evidence recovered should be labeled at the scene, so that no questions occur as to actual possession during the procedure of marking and storing the evidence. Forms detailing the processing of the evidence must be signed by both the archaeologist and police during the chain of custody.

Recovery of a buried body

The recovery of a buried body requires a preplanned methodical approach. Although the methodology to be employed is archaeological, the recovery should be processed as a crime scene. The first priority is to avoid contamination of the scene and disturbing any evidence. The area must therefore be secured, preferably by taping off the total area. A suitable access route must be established, and all people not directly involved with the recovery must not be allowed on the scene. All movement should be on the established route, which should not disturb any evidence. Record the access route in the notes on the scene and indicate it on the plan. Security is needed overnight if necessary.

Prior to excavation the scene should be photographed from various directions. Detailed notes must be taken on all the aspects of the scene such as the vegetation, climate, topography, geology, amount of sunshine, and any other observations. Flying insects should be caught at this stage, and all other insects collected and recorded where they occur in the excavation.

The surface should be cleared of vegetation without disturbing any of the evidence, after botanical samples have been taken. A three-dimensional grid system can be established with a fixed elevated datum point. If possible the grid should be oriented in north–

Table 1 Basic archaeological tools used in surface and burial site excavation.

Setting up the grid
Two tape measures (for shorter and longer distances)
Line level
Metal stakes or nails (to set up the grid)
String (to define the boundary of each trench as well as grid lines). A different color string can be used for the baseline
Compass (to determine the orientation of the buried body and grid). A prismatic compass can also be used for surveying, although more accurate equipment such as a theodolite, aladate or 'dumpy' level should be used if available
Pruning shears or lopper and a saw (to remove tree roots and branches)
Searching for and excavating the remains
Spades and pickaxes (for a deep grave)
Trowels (to remove soil)
Dental picks or bamboo skewers (for finer work around bones)
Paint brushes, big and small (to gently remove soil around bones and artifacts)
One or two buckets (to collect and carry the excavated soil to be screened)
Screens, with 1.5 mm and 5 mm mesh (to screen the soil from the burial)
Documenting
Indelible pens (to write tag names, bags or boxes)
Notebook, felt tip and ballpoint pens, pencils, erasers, ruler, graph paper, scissors
Two cameras for black-and-white and color slide photography and film
Scale, indicating at least 50 cm with centimetric divisions for detail photographs
Arrow which may be part of the scale (to point to magnetic north on photographs)
Molding agent such as plaster of Paris, silicone or dental algenate (for molding possible footprints and tool marks). Suitable containers and a spoon or spatula for mixing the molding agent
Releasing agent, such as ski wax (for use during molding)
Handling and packaging of remains and evidence
Containers and plastic bags, big and small, for insects, bones, teeth and physical evidence recovered
Packaging material, such as bubble plastic, bags, screw top bottles and boxes
Tape to seal containers
Labels
Rubber gloves and protective clothing if soft tissue is present

south direction to make reconstruction easier. A scale plan of the site should be drawn and all evidence recorded with reference to the grid.

If any footprints, tire tracks or drag marks are visible on the surface, they should be described, measured, photographed and cast. A 'dam' must be built around the evidence to keep the casting agent in place. A releasing agent such as ski wax must be applied before the casting agent is poured. The pouring should be gentle and even over the entire surface of the evidence. Make sure that the cast is strong enough to be removed without breaking after it has set. The cast can be reinforced by placing thin wire or bamboo skewers crosswise on the surface of the wet casting agent. Sometimes prints are found in a soft matrix that will not support casting in this manner. Here the print can be sprayed with a mist of cellulose acetate dissolved in acetone. As the mist settles on the surface, the acetone evaporates leaving a fine molded skin of acetate. Suitable commercial silicones can also be used for this purpose. This skin mold can then be used as the basis of a conventional cast.

If the grave was located before it was disturbed, expose the burial pit by removing the overburden with a trowel. Scraping with a trowel exposes differences in color, whereas sweeping with a soft brush shows differences in texture. Draw a plan of the surface features and topography. If the grave was disturbed by construction or other activities before it was recognized, clean all evidence *in situ* and document it before removal. If some of the surface features are still undisturbed, these should be recorded with great care to reconstruct the disturbed part of the scene.

Depending on the situation, it can be recommended that half of the feature fill should be excavated initially, without disturbing the grave pit. This can be accomplished by excavating some 5 cm away from where the walls of the shaft are expected to be and then poking at the remaining infill adhering to the wall with a sharp trowel. Usually the infill is dislodged in this manner without damaging the burial shaft. If any indications of tool marks exist in the walls of the burial pit, they should also be cast. Suitable roots should be sampled and recorded at this stage if they occur.

Excavation should follow natural strata if possible, otherwise it should be in horizontal arbitrary layers of

10–15 cm. If natural stratigraphy does occur, the excavation layers should not exceed 15 cm. Natural strata thicker than 15 cm should be divided in excavation spits. Collect a number of soil samples from the grave fill above the remains. Usually a soil sample from every spit is sufficient. Soil from the excavation should be screened.

All evidence found in the infill should be left *in situ*, or if impossible it should be thoroughly documented before removal. As soon as there are indications that the remains have been located, excavation should be stopped to record the profile of the section and the burial pit. The remains should be left covered with as much soil as possible in order to prevent damage to the bones during the rest of the excavation.

When all the remains have been located and the features of the grave documented, the rest of the grave fill can be removed. Excavation should always progress from the center towards the edges to avoid repeated brushing around bones which can dislodge them from their position. Burials in a forensic context are seldom very deep. In a deep grave, the excavation should be expanded if the remains have not been found at a depth where they can easily be reached from the surface. Usually more damage is done when the edge of the grave pit is collapsed by an excavator lying on the surface and reaching into the grave than by a controlled extension to the excavation. The extension should be planned so as to cause minimal damage to the walls of the burial pit, and all features of the pit should be recorded before it is sacrificed. The floor of the extension or trench must be deeper than the level on which the remains occur in the grave in order to avoid contaminating the burial pit. By extending the excavation a work area is created bringing the excavator within easy reach of the remains. A trench roughly in the middle of the grave, and wide enough for one person to sit in, is usually the best option (**Fig. 3**).

Once the remains are visible, the bones are exposed in the top-down manner and from the center to the edges. Again, all bones and associated physical evidence should remain *in situ* until documented and photographed (**Fig. 4**). A clear description of the orientation of the body and the burial position is recorded, drawn and photographed and soil samples from the abdominal cavity collected. All bones should be removed with care and securely packed after documentation. A test excavation conducted immediately below the level at which the remains were located will ensure that all evidence has been recorded and removed. The profiles and vertical shape of the grave pit should be recorded before the excavated area is refilled.

The same procedure can be used for intact and less-

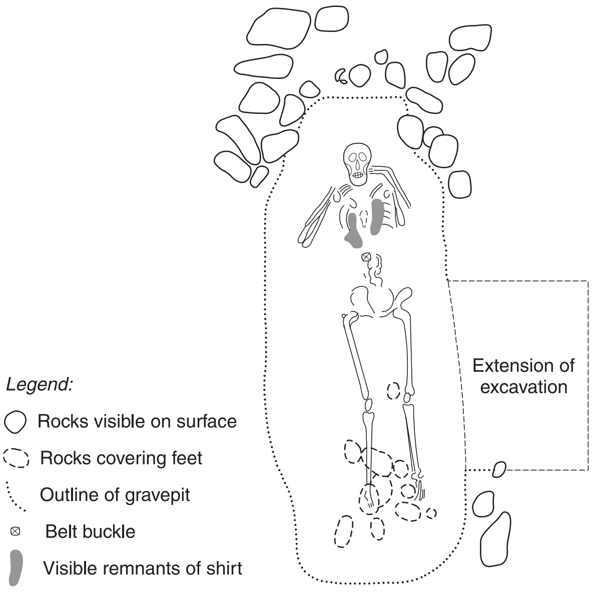


Figure 3 Extension of excavation to create work space for excavators.

Figure 4 Exposed human remains *in situ*. Note arrow, scale, date and other particulars indicated. 1–10, Detail indicated on photograph.

decomposed remains and in situations such as mass murder. These remains are always highly odorous. Excavators should take care to protect themselves by wearing gloves and protective clothing. The corpse should be put in a tagged body bag immediately.

Recovery of surface scatters

Not all remains are buried; in fact, most of the forensic remains are found on the surface. They may be scattered if the deceased wore summer clothing or might be better preserved in a winter outfit. In the case of scattered remains, the area should be walked systematically and carefully without stepping on any evidence. Every object should be flagged, but not removed before documentation. Care should be taken not to miss smaller objects such as a ring or tooth. Soil around larger pieces of bone can be screened after they have been removed.

When all possible pieces are located, the source from where the remains were scattered should be determined and the agent responsible for the scattering identified. In the case of dogs, for example, a knowledge of their habits may lead to the discovery of more remains. Ribs are very often dragged away by scavenging animals. Skulls may roll away if the body is on an elevated surface and therefore the lower reaches of the slope should be searched. Determining the path and direction of flow of water on a slope or in a valley may also lead to the discovery of more remains. The position of evidence and remains must be surveyed and indicated on a scale map of the locality (**Fig. 5**). They should also be labeled and their orientation noted. The same procedures used in buried body recovery apply in so far as the grid and documentation of the site is concerned. The area must also be tested with a metal detector to see if any metal objects (e.g. bullets) are on the surface or in the ground. As in the burial excavation, soil under the remains should be screened to locate smaller artifacts and skeletal fragments in case they are not visible.

The same kind of approach can be utilized for burned remains and mass disasters, but each case should, of course, be evaluated individually and the techniques adjusted accordingly.

Documentation

One of the important principles in any excavation is the fact that it is destructive. Therefore, care should be taken to record all stages of the excavation by means of complete written notes, photographs and drawings. Photographs should be taken in both color and black-and-white. Each photograph should show the magnetic north and the scale. In practice, a small magnetic or blackboard works very well, on which

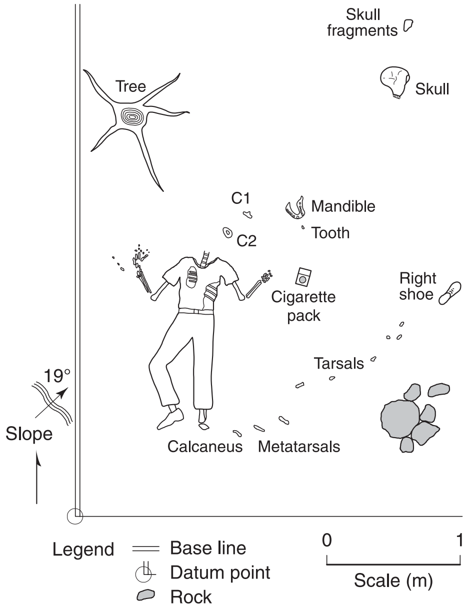


Figure 5 Sketch plan of a surface scatter. Note scale, north arrow and datum point.

the date, location, case number, orientation and depth can be indicated (see Fig. 4).

Specific observation must be made of whether the bones are articulated, which indicates that the body was most probably still intact when buried. The position of all limbs must be noted. Photographs taken should include images of the skeleton as a whole, as well as close-ups of any special finds. These finds are specifically marked to make them more visible on the photographs (see Fig. 4). Detailed notes which include the depth of the remains should be taken, and a scale diagram drawn of the grave from above and in cross-section. The depth below surface and the stratigraphy involved are also very important, and it may be necessary to draw the four profiles (east, south, west and north walls) of the grave.

Sampling at the scene

The matrix of the burial forms part of the evidence at the scene. The soil could have been mixed with chemicals, or pollen from the soil might be specific enough to trace the origin if the body had been transported over a long distance before it was buried. Trace evidence of poison and narcotics may also be preserved in the matrix. The only way to determine if these are present is by laboratory analysis of soil samples.

Plants occurring on the surface of the grave and in the surrounding area should be collected and pressed in a conventional plant press for analysis by a forensic botanist. It is sometimes possible to determine the

postmortem interval by calculating the time it would take for a specific plant species to grow to the size of the collected specimens. Roots that were cut while the grave was being dug should also be collected from the profiles of the grave. By studying the ends and growth rings the forensic botanist can sometimes determine the season in which the root was damaged. Roots growing into the infill may also indicate the time that has elapsed since the grave was dug.

Insects, larvae and pupal cases should be collected and their location recorded. Insects flying in the vicinity of the body must also be collected. Live insects and larvae can be placed in small vials with a piece of damp cotton or tissue paper. It is necessary to collect live insects because it is sometimes difficult to determine the species from larvae. The live juveniles are hatched and the adult can then be classified. Insects should also be preserved at the time of collection to record the stage of development. This is done by placing them in vials filled with a suitable preservative. It is important to note where the insect was collected. An entomologist should be consulted to analyze the specimens.

Sampling for DNA analysis may be also necessary. Any bone that is enclosed by a volume of cortex, such as the body of a vertebra or a foot bone, may be suitable for this purpose. It is important to sample a bone that protects a part of itself from contamination and is also not needed for anthropological analysis. Samples for DNA analysis should be collected while wearing gloves, and bones that have been handled by the excavators should be avoided. DNA can also be obtained from teeth or cancellous bone such as the head of the femur, but this should only be done after the osteological analysis has been completed. The best samples, however, are those collected at the site to eliminate the problem of contamination.

Discussion

Most of the procedures described are based on meticulous and systematic observation which should lead to the maximum amount of information and material retrieval. It should be re-emphasized that once the skeleton or an artifact has been removed, it can never be placed back into its original position. The most important aspect of the excavation is to provide a lead for the medicolegal investigators and police in order to identify the victim. Another aspect of such an excavation is to testify in court if necessary, and therefore the documentation must be complete and the chain of custody insured. Both modern and archaeological graves are valuable sources of information for forensic anthropologists. Repeated documentation, with the focus on context, by various means, e.g. photographs, videos and written descriptions, are crucial at all stages of the operation.

The chain of evidence should be maintained at all times, and the person in charge of the case must be aware of the whereabouts of all evidence including the remains at all times. All evidence should be signed for at all stages when they are handed over to the care of another person.

The advantages of systematic, scientific methods of excavation and documentation are obvious. If done properly, relevant information will be available permanently in case there is a need to reopen the case after some years. A scientific approach provides great accuracy in the collection of evidence and a higher probability that all physical evidence is collected, prevents postmortem damage to bones and artifacts, and decreases the probability of destroying contextual information.

See also: **Anthropology**: Determination of Racial Affinity; Morphological Age Estimation; Skeletal Trauma; Overview; Archaeology. **Crime-scene Investigation and Examination**: Recording; Recovery of Human Remains; Packaging; Preservation; Scene Analysis and Reconstruction.

Further Reading

Boddington A, Garland AN and Janaway RC (eds) (1987) *Death, Decay and Reconstruction: Approaches to Archaeology and Forensic Science*. Oxford: Manchester University Press.

Dirkmaat D and Adovasio JM (1997) The role of archaeology in the recovery and interpretation of human remains from an outdoor forensic setting. In: Haglund WD and Sorg MH (eds) *Forensic Taphonomy: the Postmortem Fate of Human Remains*, pp. 39–64. Boca Raton, FL: CRC Press.

Erzinçlioglu YZ (1986) Areas of research in forensic entomology. *Medicine, Science and the Law* **26**:273–278.

Galloway A and Snodgrass J (1998) Biological and chemical hazards of forensic skeletal analysis. *Journal of Forensic Sciences* **43**:940–948.

Hunter J, Roberts C and Martin A (eds) (1996) *Studies in Crime: An Introduction to Forensic Archaeology*. London: Routledge.

Killam EW (1990) *The Detection of Human Remains*. Springfield, IL: Charles C. Thomas.

Krogman WM and İşcan MY (1986) *The Human Skeleton in Forensic Medicine*. Springfield, IL: Charles C. Thomas.

McLaughlin JE (1974) *The Detection of Buried Bodies*. Yuba City, CA: Andermac.

Melbye J and Jimenez SB (1997) Chain of custody from the field to the courtroom. In: Haglund WD and Sorg MH (eds) *Forensic Taphonomy: the Postmortem Fate of Human Remains*, pp. 65–75. Boca Raton, FL: CRC Press.

Miller PS (1996) Disturbances in the soil: finding buried bodies and other evidence using ground penetrating radar. *Journal of Forensic Sciences* **41**:648–652.
Morse D, Crusoe D and Smith HG (1976) Forensic archaeology. *Journal of Forensic Sciences* **21**:323–332.
Owsley DW (1995) Techniques for locating burials, with emphasis on the probe. *Journal of Forensic Sciences* **40**:735–740.
Quatrehomme G, Lacoste A, Bailet P, Grevin G and Ollier A (1997) Contribution of microscopic plant anatomy to postmortem bone dating. *Journal of Forensic Sciences* **42**:140–143.
Skinner M (1987) Planning the archaeological recovery of evidence from recent mass graves. *Forensic Science International* **34**:267–287.
Skinner M and Lazenby RA (1983) *Found! Human Remains: A Field Manual for the Recovery of the Recent Human Skeleton*. Burnaby, BC: Simon Fraser University Archaeology Press.

Morphological Age Estimation

S R Loth, Department of Anatomy, Faculty of Medicine, University of Pretoria, Pretoria, South Africa
M Y İşcan, Adli Tıp Enstitusu PK10, Cerrahpasa, Istanbul, Turkey

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0755

Introduction

When human remains are rendered unrecognizable by advanced decomposition or are completely skeletonized, experts must be able to make the four determinations: age, sex, race and stature. This information is the most essential aspect of the identification process and indeed, gives police a starting point for their investigation. Of these demographic characteristics, the estimation of age at death is the most challenging once adulthood had been reached. As in the living or recently deceased, precise estimation of age without documentation such as a birth certificate or driver's license is not a simple task because of the tremendous variability in the manifestations of the aging process. How humans look as they age is orchestrated by the complicated interplay of genetic, environmental, and cultural factors. Everyone's hair does not turn gray at the same age or rate. The same goes for the development of wrinkles and other features associated with advancing age. Moreover, a face lift and hair dye can make someone look much younger, whereas smoking and sunbathing will add years. Even though plastic surgery does not alter the appearance of the sites that osteologists depend on for estimating age at death, the skeleton also shows the effects of age, and these are just as subject to variation.

There are, however, some basic biological processes that all humans undergo from the prenatal months through infancy and childhood and finally to adulthood. If one understands the growth process, it becomes possible to separate early childhood from 1 to 12 years and adolescence from 13 to 18 years. Adulthood can also be divided into young, middle and older stages. The events of these periods are also reflected in the skeleton. For example, early childhood is characterized by dental growth and development, adolescence by completion of bone growth, middle age by remodeling and maintenance that gradually decline to the wear and tear of old age.

The aim of this article is to present some of the basic dental and skeletal changes that can be used as age markers in an examination of unknown remains. The literature on age estimation can be found in several books and leading review articles.

Dental and Skeletal Growth

Bone growth and dental development have been studied by skeletal biologists interested in the skeletal remains of children in both prehistoric and modern populations. During the early years, there are basically two major physiological activities in progress: dental development and eruption, and bone mineralization and ossification, both of which hold important clues to the age of an immature person. Under normal circumstances, the order of events during the developmental stages follow regular, predictable patterns with little variability.

The most reliable way to estimate age in children is based on the sequence of both deciduous and permanent tooth formation and eruption. These are excellent age markers because there is consistency in the timing and sequence of eruption. Of the dental sets, the deciduous dentition is first visible at about 6–7 months and is completed by 3 years (**Table 1**). During this time, eruption begins with the central incisors in the mandible and ends with the second molar. Although the sequence is well established and rarely varies, population variation exists in the timing of dental eruption and this has been studied in many different parts of the world. **Tables** 1 and **2** display ages of eruption for deciduous and adult dentition, respectively, in several populations. Some researchers claim that accuracy can be best estimated from the enamel formation in the deciduous dentition (up to

Table 1 Age of eruption (in months) for deciduous teeth

Population	*i1*	*i2*	*c*	*m1*	*m2*
Maxilla					
Egypt	9.4	11.2	19.1	14.1	24.7
Tunisia	9.3	12.5	18.3	15.7	22.0
USA	7.5	8.0	18.0	14.0	25.0
Canada	9.0	10.2	18.0	15.2	27.5
New Guinea	10.6	11.9	18.2	16.0	24.1
Mandible					
Egypt	8.3	12.2	19.1	14.1	24.7
Tunisia	6.0	14.0	19.0	15.7	21.7
USA	6.5	7.0	18.0	14.0	25.0
Canada	7.1	12.0	18.2	15.3	26.5
New Guinea	8.2	13.0	18.8	16.6	23.9

Modified from Novotný *et al.* (1993).

age 4 years) and in the ensuing permanent teeth as well as the root.

Figure 1 contains a sketch of the development of both deciduous and permanent teeth. It indicates the level of tooth development and time of emergence at a given age. As can be seen in this figure, M1 is the first permanent tooth to erupt. Observations of the roots reveal that the canine and M2 are the last of the regular sequence to be completed. In general M3 is not used in age estimation because of the enormous variation in the timing of its eruption. One must also bear in mind that there is a continuing microevolutionary reduction in the dental apparatus a major feature of which is the trend toward impaction and congenital absence of 'wisdom teeth', especially in Whites. It should be noted that Fig. 1 gives general age ranges for the stages of development rather than precise times for these events and does not account for sex and race differences. All research on this topic has clearly demonstrated that the sequence of development and eruption begins earlier in females and Blacks. There is no question that the dental sequence is the most reliable and accurate means of age assessment during the first 12 years of life.

Table 2 Age of eruption (in years) of permanent dentition

	Upper			*Lower*		*Upper*
Country	*I1*	*I2*	*C*	*P1*	*P2*	*M2*
Boys						
Czech	7.2	8.1	11.2	10.4	11.0	12.1
Egypt	8.0	9.1	12.1	11.3	12.1	12.7
England	8.1	8.9	12.4	11.4	12.1	12.4
Hungary	7.3	8.5	11.4	11.8	11.9	12.4
USA	7.5	8.7	11.7	10.8	11.4	12.7
Japan	8.0	9.2	11.6	10.8	11.6	12.9
Girls						
Czech	6.9	7.7	10.6	9.4	10.6	11.7
Egypt	7.3	8.6	11.8	10.8	11.2	11.7
England	7.7	8.7	11.0	11.4	12.0	12.1
Hungary	7.1	8.0	11.6	10.0	11.3	12.0
USA	7.2	8.2	11.0	10.2	10.9	12.3

Modified from Novotný *et al.* (1993)

Bone formation and growth can also be used as an age indicator in the immature skeleton. The formation of bone tissue (ossification) proceeds in a systematic and organized manner. Beginning at loci referred to as centers of ossification, development normally follows a well-known sequence and timing of events culminating in the development of the diaphysis and epiphysis. Yet, as is the case with most human biological functions, there can be variability in all aspects of this process, primarily in timing, and to a lesser extent, in sequence. During adolescence, the cartilaginous growing region or metaphysis of the endochondral bones is replaced by bone to unite the epiphyses with the diaphyses.

Much of our current knowledge of bone formation and development has been derived from radiological studies, and age-related skeletal changes have been studied from the fetal stage through late adolescence. Prenatally, age estimation is based on size differences in each of the many fetal bones. Growth proceeds at its greatest velocity during intrauterine life and infancy, but slows considerably once the 2nd year is attained. The next major growth spurt occurs at puberty accompanied by the development of secondary sex characteristics.

Fontanelle closure in the infant skull is another source of clues to age at death. The posterior fontanelle begins ossifying at 2–3 months after birth and closes by about 6 months, whereas the anterior fontanelle shrinks after birth but does not close completely until about 2 years of age. The metopic suture which runs from nasion to bregma on the frontal bone normally closes by around the age of 6 years. In rare instances, it remains open throughout life.

As with the dentition, bone growth is also subject to variability by sex and population. Radiological growth charts are available for several populations to assess age of epiphyseal growth and union. However, there are very few direct studies of skeletal remains because of the scarcity of children in documented collections. On average, epiphyseal fusion begins at about 12–14 years. This process starts earlier in females as a component of their earlier attainment of sexual maturity. The first joint of the body to unite is the elbow formed by the articulations of the distal humerus and proximal radius and ulna. Fusion then proceeds in a set order throughout the skeleton

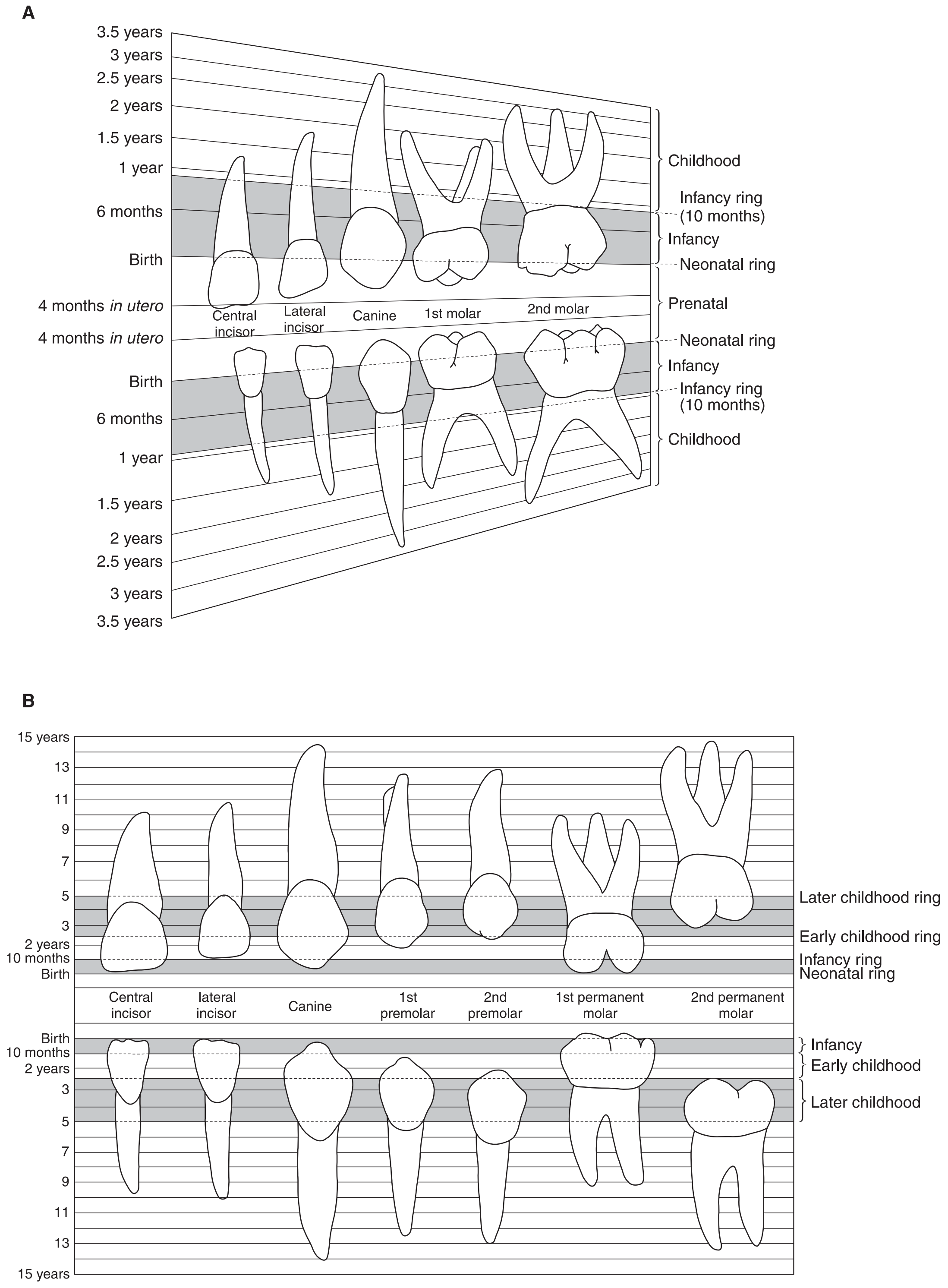


Figure 1 Ages of development and eruption of deciduous teeth (**A**) and permanent dentition (**B**). (Modified from Massler *et al*. (1941) Developmental pattern of the child as reflected in the calcification pattern of teeth. *American Journal of Diseases of Children* **62**: 33.)

until growth is complete sometime between the late teens and early 20s with males active at the older ages. **Figure 2** is a simplified sketch showing the order of epiphyseal fusion beginning first in the elbow and ending in the shoulder. In general, there is about a one year differential between each of the joints; thus it normally takes about 6 years to complete epiphyseal union ranging from about 12–14 years to approximately 18–20 years. Again, sex and population differences must be taken into account. By the early 20s, most of the epiphyses are united with the exception of the basisphenoid synchondrosis of the skull which has been observed to close as late as 29 years and the medial clavicle which may remain open into the 30s. Although a long bone length-based age estimation has been suggested, growth in adolescence is so much more varied than that characterizing the intrauterine months and early childhood, that epiphyseal union is by far the most reliable age marker in the teens. The best estimates are obtained if figures have been drawn from the same population as that of the skeleton in question.

Adult Age

Unlike the regular predictable order of growth and development in childhood, once maturity is attained, skeletal structures are modified and maintained by the process of remodeling. This activity can be highly variable from one individual to the next and the alterations produced are usually subtle and often difficult to detect and interpret. The adult aging process is susceptible to the influences of human behavior and the environment. This is clearly seen in the fact that even genetically identical twins can differ from each other as they get older. Thus, age estimation in the adult is one of the most challenging assessments to make from the skeleton, and middle age is probably the most difficult because it is an

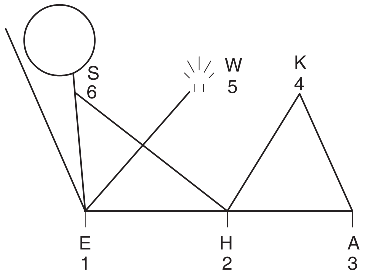


Figure 2 Order of epiphyseal closure beginning with the elbow joint (E) at about 12–14 years of age, followed (at approximately 1 year intervals) by the hip (H), ankle (A), knee (K), wrist (W), and ending with the shoulder (S) by 18–22 years. (Modified from Novotný *et al.* (1993)

extremely varied transitional period hormonally and metabolically. The later years also have their share of complications, including inevitably higher rates of pathologic conditions and the effects of a lifetime of wear and tear. Of particular concern to the forensic anthropologist is the fact that few skeletal components show discernible patterns of age-related change past the 40s.

The first attempts at the development of adult age estimation techniques from systematic studies of (supposedly) known aged skeletons began in the 1920s with observations of the pubic symphysis and cranial sutures. These were followed by years of attempts to improve accuracy at these sites. Unhappily, cranial suture closure is so inherently erratic that it remains a technique of last resort and only used when no other alternative is possible. For many years and with many modifications, the pubic symphysis had been considered more reliable, but now experts have conceded that it has a limited range of usefulness. It is problematic by the 30s and ineffective by the 40s. Nearly every part of the skeleton has been examined for adult age assessment, but few sites have proven to be consistently reliable. Different methodological approaches such as radiography and histology have also been attempted, but with limited success.

It took over 60 years until serious attention was shifted to any other part of the skeleton than the cranium and pubic symphysis. Today the most consistently reliable method for the determination of age at death from the adult skeleton is based on a lifetime of changes at the sternal end of the rib that have been divided into a series of phases beginning in the late teens. Standards for white males and females have been extensively tested with consistently accurate results on many diverse modern and archaeologic populations. **Table 3** shows that accuracy ranges from as little as ± 1.5 years through the 20s to about ± 5 years thereafter. The youngest phase (0) and oldest (8) are open ended (e.g. less than 17 or over 70).

The aging process in the rib was assessed by studying changes at the costochondral junction of the fourth rib (**Fig. 3**). The metamorphosis of the rib begins after the completion of growth at the sternal extremity at about 14 years in white females and 17 in their male counterparts. This is consistent with the generally earlier maturation of females. It is characterized by the disappearance of the epiphyseal line and the beginning of pit formation at the nearly flat, billowy surface of the sternal end of the rib. Within a few years, the pit has taken a definite V-shape and scallops appear on the rim of the bone. The pit gradually deepens and widens to a wide V (in females)

Table 3 Comparison of mean ages and 95% intervals for rib and pubic symphysis phases

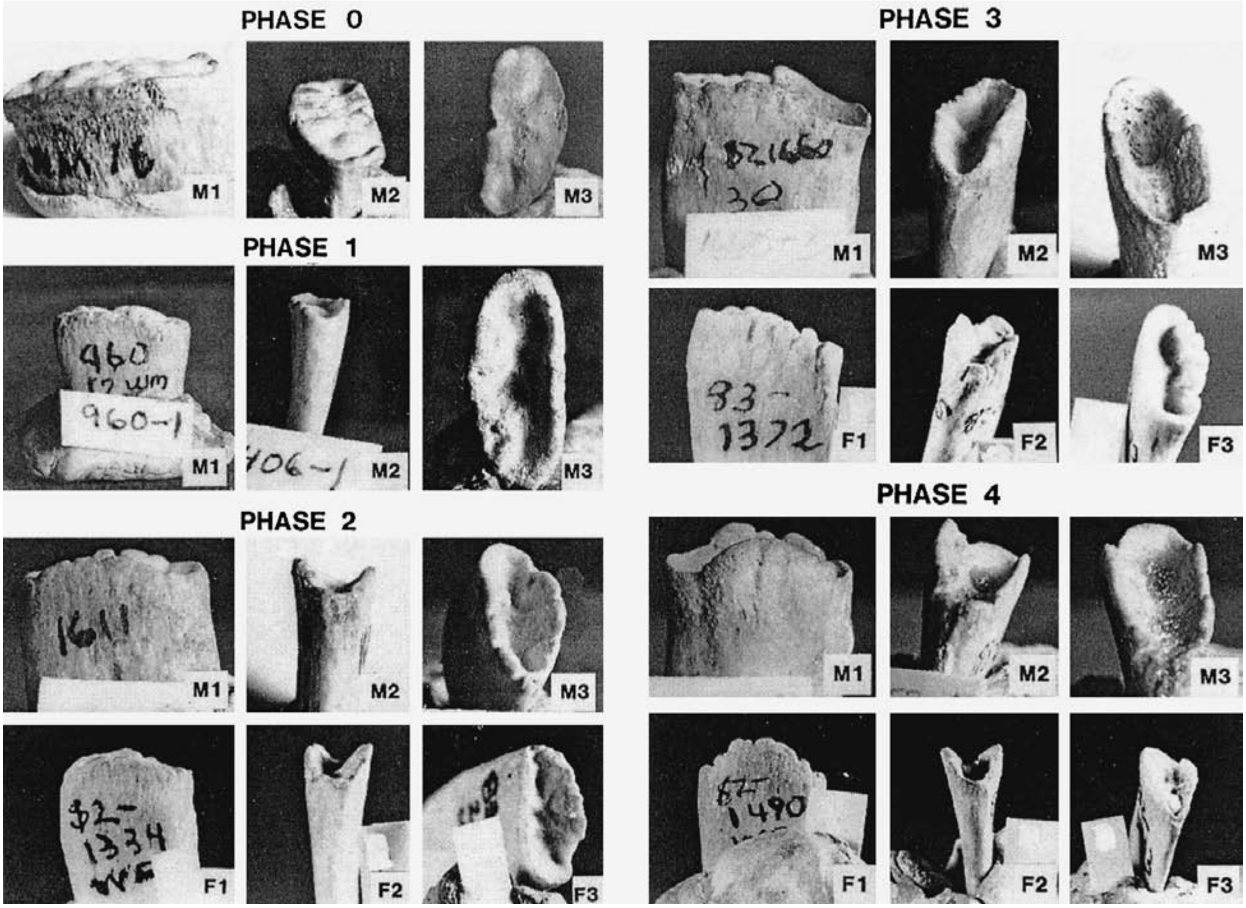
	Rib		*Pubic symphysis*	
Phase	*Mean*	*95% Interval*	*Mean*	*95% Interval*
Males				
1	17.3	16.5–18.0	18.5	15–23
2	21.9	20.8–23.1	23.4	19–34
3	25.9	24.1–27.7	28.7	21–46
4	28.2	25.7–30.6	35.2	23–57
5	38.8	34.4–42.3	45.6	27–66
6	50.0	44.3–55.7	61.2	34–86[a]
7	59.2	54.3–64.1		
8	71.5	65.0–[a]		
Females				
1	14.0			15–24
2	17.4	15.5–19.3	25.0	19–40
3	22.6	20.5–24.7	30.7	21–53
4	27.7	24.4–31.0	38.2	26–70
5	40.0	33.7–46.3	48.1	25–83
6	50.7	43.3–58.1	60.0	42–87[a]
7	65.2	59.2–71.2		
8	76.4	70.4–[a]		

[a] Terminal phase age ranges are open ended. Modified from Loth and İşcan (1994)

or U shape in both sexes. The rounded edges of youth begin to thin and sharpen by the mid 30s. With increasing age, the rim becomes irregular, the interior of the pit becomes porous and the bone quality deteriorates until the ribs of most individuals over 70 years are thin and irregular with bony projection at the costochondral junction.

There are considerable race and sex differences in bone density with age. Bone density decreases dramatically in Whites, and especially females, whereas it remains minimally changed in Blacks, with the least deterioration in males. However, the same process that contributes to the maintenance of density over age 30 (periosteal deposition) also makes the rib look

Figure 3 Rib cage phases 0–8: metamorphosis of the sternal end of the rib in males (M) and females (F). The solid, smooth immature rib has a flat or billowy end with rounded edges bordered by the epiphyseal line (phase 0). This site undergoes a series of changes characterized by the formation and deepening of a pit at the costochondral junction, accompanied by thinning and porosity of the bone and development of sharp, irregular edges throughout the life span to Phase 8 in extreme old age (over 70 years) (see Table 4). (Modified from İşcan and Loth (1989) The osteological manifestations of age in the adult. In İşcan MY and Kennedy KAR (eds) *Reconstructions of Life from the Skeleton*, pp. 23–40. New York: Wiley–Liss.



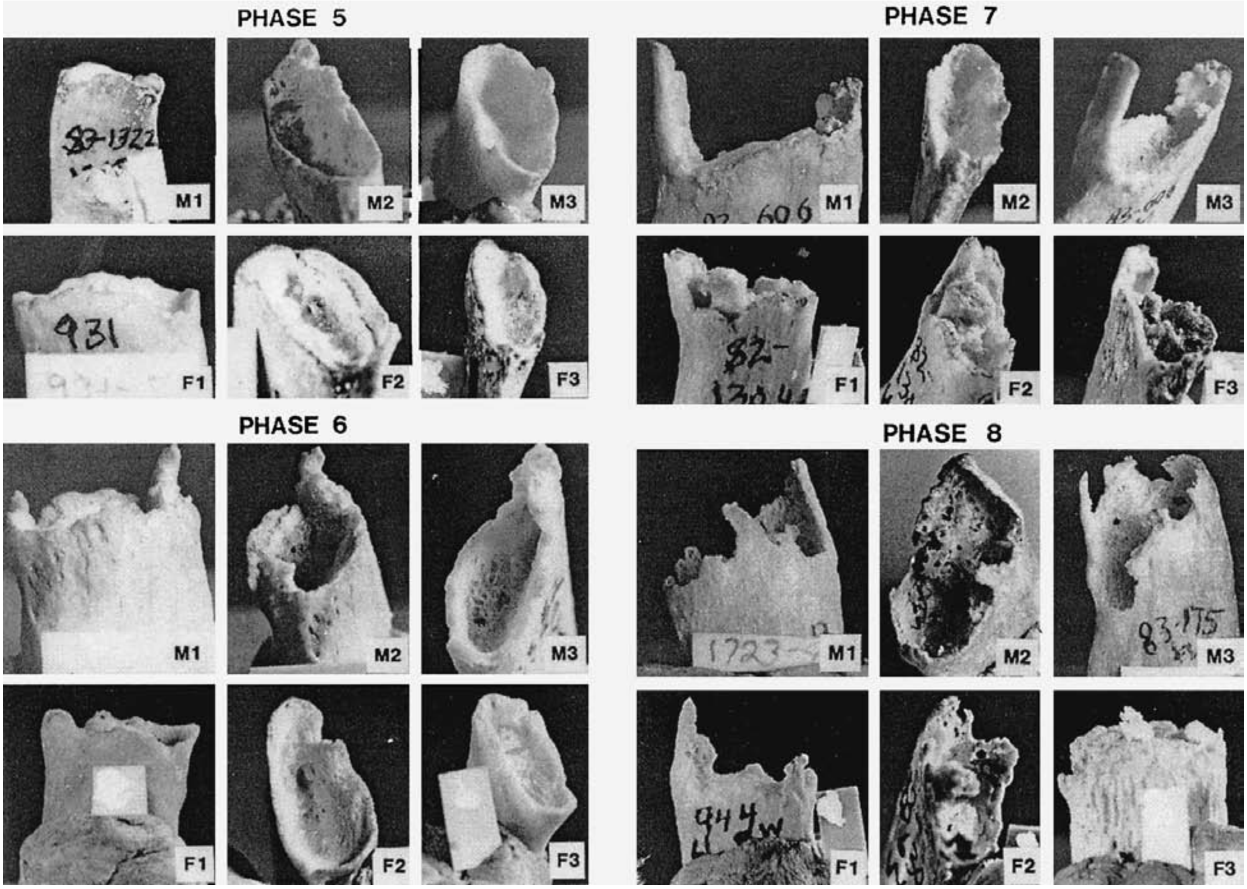
older because it creates irregular bony projections in Blacks (particularly males) decades before these features appear in Whites. Concerns over side and intercostal differences have been pretty much dispelled. Several studies have shown that although standards are based on the right fourth rib, there are no side differences and intercostal variation among true ribs 2–7 and especially between ribs 3, 4 and 5 is not significant.

One reason the rib phases are quite precise and reliable is the fact that sex and race differences were investigated and accounted for in the standards. Since, as noted earlier, these factors produce significant variation in the rate and pattern of rib metamorphosis through the lifespan. Moreover, there is little individual variation because the only direct stress on the ribs is breathing. This contrasts with other skeletal sites like those in the pelvis which are directly involved with weight bearing, pregnancy and parturition.

Although the rib is the bone of choice, the sternal end is not always present or intact so the forensic anthropologist must be prepared to evaluate whatever information is available from other skeletal components. In addition, the odds of correct age estimation increase if other sites yield similar results.

Before the development of the rib phase method, the pubic symphysis was considered the most reliable morphological technique. The general aging pattern at this site is represented in **Fig. 4**. Standards have been reworked numerous times to attempt to cope with great variability, and casts have been created to help improve the effectiveness of matching the symphysis to an age phase and reduce interobserver error. However, the results of independent tests of several pubic symphyseal techniques were found to be disappointing in accuracy and precision. Furthermore, as can be seen in Table 3, even though a bone can be matched more easily to a cast, the 95% confidence intervals of the age range is too large to be very useful for most phases.

The latest pelvic site to be tapped is the auricular surface of the ilium. Developed for use on archaeologic samples where it was possible to seriate a large number of specimens, the workers in this area have reported good results on early cemetery and American Indian burials. However, this method is very difficult to apply and independent testing indicated that this site was not suitable for use in individual forensic cases.



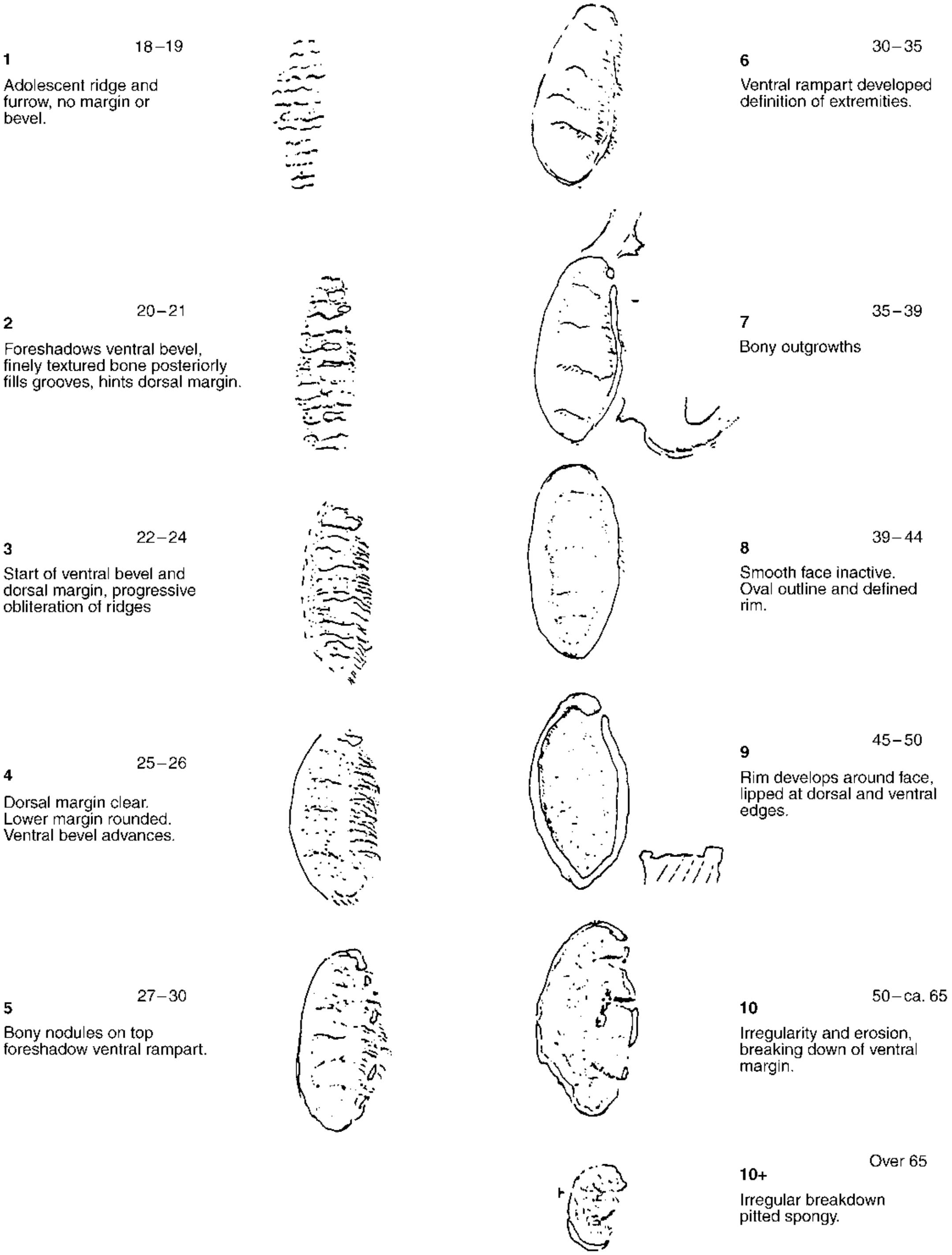


Figure 4 Schematic of pubic symphyseal changes with age.

Often only a skull is found or made available to the forensic anthropologist for analysis. Unfortunately, there are no precise age indicators in this part of the skeleton and one can only give a very rough estimate such as young adult, middle aged or elderly. In the skull, the cranial suture closure pattern and its relation to age has been rigorously investigated for nearly a century. In spite of the considerable attention paid to this technique, variation is so enormous that it is practically abandoned as an age marker. It is not all that rare to see an entirely fused skull suture in the 20s or patently open ones in the elderly. It has been suggested that because a suture is a long structure it can exhibit a range of intrasutural variability along its

length. In an attempt to reduce this problem, the most recent method offered a modified approach. Instead of examining the entire suture, only 1 cm sections were selected for assessment at specified locations. The closure pattern at each spot was scored as follows:

Score 0: Completely open. No ectocranial closure.
Score 1: Minimal to moderate closure, from a single bony bridge to about 50% synostosis.
Score 2: Significant, but not complete closure.
Score 3: Complete obliteration and fusion.

The five lateral-anterior suture sites (1–5 in **Fig. 5**) yielded the best results and the total score can be converted into an age estimate using **Table 4**. The mean ages using this technique ranged from 32 to 56 years. It must be emphasized, however, that the developers of these standards note that there is still great variability and recommend that this site be used in conjunction with other indicators. Attempts have been made to correlate age with the closure of other sutures of the vault and facial skeleton, but results have not been satisfactory.

The association between tooth wear and aging has long been known and is still commonly used for archaeologic assemblages. This process has been studied by many scientists for nearly 100 years. It is a generally reliable indicator in relatively small, ancient populations that were genetically and nutritionally homogeneous for whom group specific rates of attrition could be calibrated from juvenile samples.

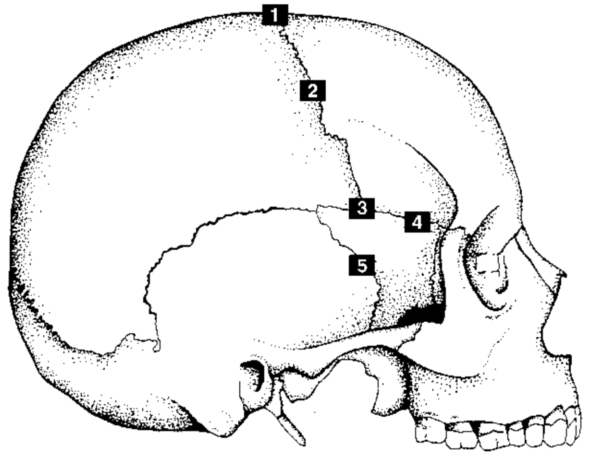


Figure 5 Suture closure at lateral–anterior sites (taken at points indicated by numbers). 1. Bregma; 2, midcoronal (midpoint of each half of coronal suture at pars complicata); 3, pterion; 4, sphenofrontal (midpoint); 5, sup. sphenotemporal 2 cm below juncture with parietal bone. (Modified from Meindl and Lovejoy (1985), Ectocranial suture closure: a revised method for the determination of skeletal age at death based on the lateral–anterior sutures. *American Journal of Physical Anthropology* **68**:57–66.)

Table 4 Age estimates from lateral–anterior suture composite scores

Composite	*Mean*		*Interdecile*	
Score	*N*	*Age*	*SD*	*Range*
0 (open)	42			x–43
1	18	32.0	8.3	21–42
2	18	36.2	6.2	29–44
3–5	56	41.1	10.0	28–52
6	17	43.4	10.7	30–54
7–8	31	45.5	8.9	35–57
9–10	29	51.9	12.5	39–69
11–14	24	56.2	8.5	49–65
15 (closed)	1			

Modified from Meindl and Lovejoy (1985).

The great individual variation of contemporary humans reduces the effectiveness of this site for aging. However, one of the most recent examples of efforts to use dental attrition for clues to age today appears in **Fig. 6**. Although the standards were developed from archaeologic Amerindians, application to turn of the twentieth century American Whites and Blacks produced surprisingly good results. The figure depicts discernible stages from A through I (in both jaws), that begin as early as 12 years and continue as late as 55 years and reports no sex differences. Cross-hatching signifies the transitional stage from the enamel to the secondary dentine formation. The black areas indicate that the enamel is entirely gone and dentine is visible. It is clear that the wear pattern goes from the anterior to the posterior teeth.

Tooth wear is likely to have the best relationship to broad or general age categories of any other indicator in the skull, but its application to forensic cases must be interpreted cautiously because (as noted above) of the tremendous individual variation in current large, heterogeneous, modern populations. These differences encompass many factors including relative enamel thicknesses, significant racial dimorphism, widely varied diet, socioeconomic status, and access to health and dental care. A further consideration is that, although the people of today live considerably longer than their archaeologic counterparts, wear rarely progresses to the extreme stages seen in the figure. This is primarily due to our relatively soft refined diet, free from the grit that literally ‘sanded away’ the entire tooth crown. In addition, tooth loss, resulting from decay and periodontal disease, is much more common now, and more so in Whites than Blacks.

Another, more comprehensive approach to dental aging in the adult was first introduced nearly 50 years

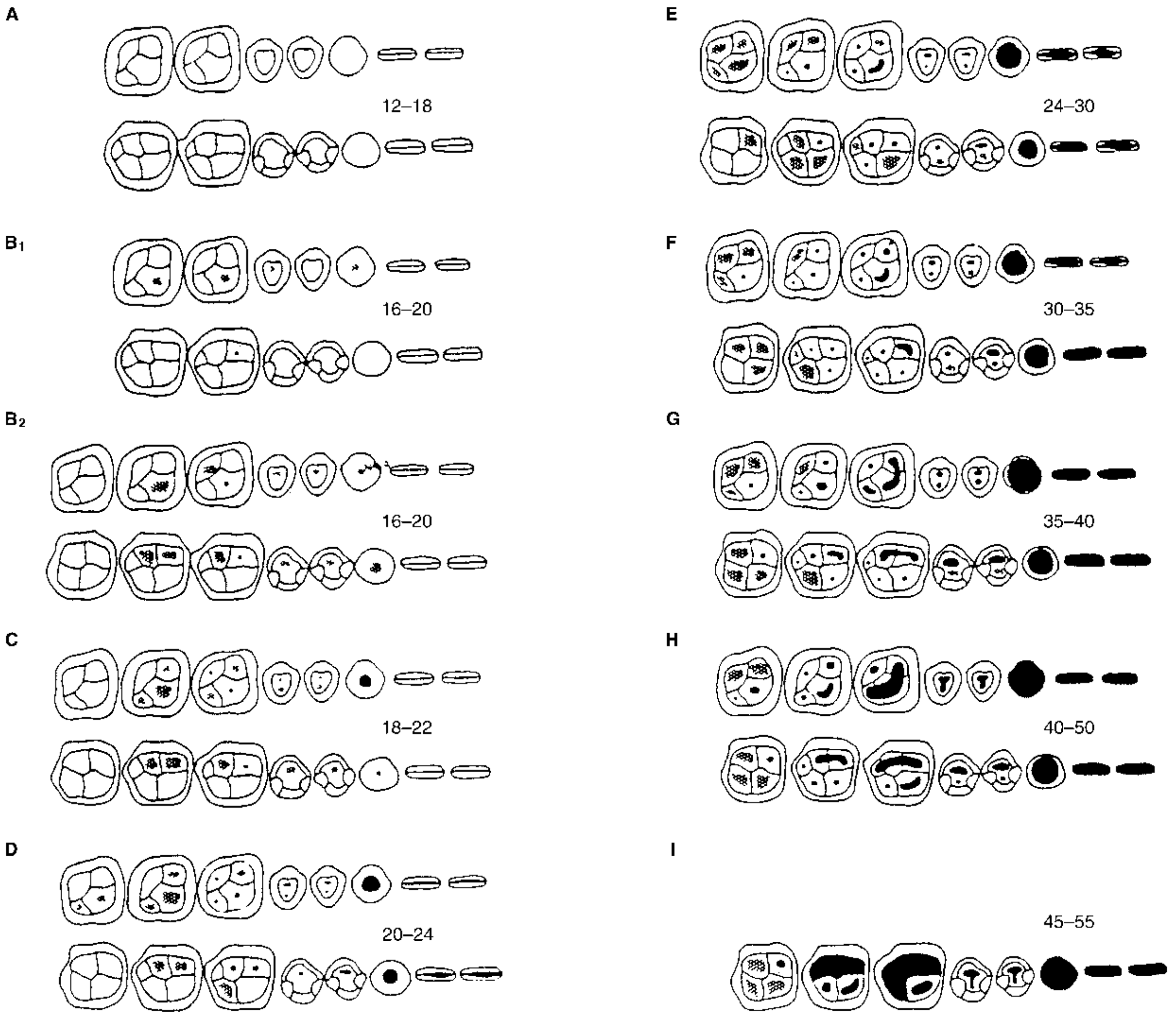


Figure 6 Age phases of dental wear in the maxilla and mandible (see Table 4). (From Lovejoy (1985), Dental wear in the Libben population: its functional pattern and role in the determination of adult skeletal age at death. *American Journal of Physical Anthropology* **68**:47–56.)

ago. It was designed to assess a number of characteristics of a cross-section of a tooth including attrition, periodontosis, secondary dentine, cementum apposition and root resorption. Since the original study there have been many modifications of this technique. Although the method can yield a reasonable age range estimate, its disadvantages include the removal and destruction of a tooth (usually a canine or premolar), and the laboratory equipment and time necessary for thin sectioning and microscopic analysis.

Other parts of the adult skeleton can sometimes be used to assign an upper or lower value. In the absence of trauma or disease, the development of osteophytic lipping on the bodies of the vertebrae is a good indication that an individual is over the age of 40 years. Bone quality can also be a guide. Firm solid bones are consistent with youth whereas thin, porous, fragile bones signal old age.

Radiologic and Histomorphometric Techniques

Radiology is not commonly used by anthropologists and few real standards are available. Some of the earlier studies go back to the 1950s when epiphyseal involution was observed. Currently, the proximal epiphysis of the humerus, femur and tibia are all well documented. Again as in many other age estimation techniques the deterioration of the spongy tissue has been sequenced and divided into several stages. Several other bones such as the clavicle and ossification of costal cartilages have also been observed. Radiography may be effective at giving a general idea of age from a mature long bone where there is no other noninvasive alternative, but it cannot be relied upon for precise estimates.

Age estimation using histomorphometry was first

experimented with in the 1960s. This technique attempts to correlate age with the cellular activity involved with bone remodeling at the histologic level by counting osteons, assorting them into different stages of development or breakdown, and calculating their ratios. The remodeling process has been studied by many investigators who found it to vary by race and sex. Population specific standards have been introduced for both contemporary and prehistoric populations. The age range covers the adult stage of human life span. Accuracy, as measured by the standard error of estimate claimed in some publications on the subject varies by around 6 years. Quantified comparisons of rib phase assessment with histomorphometric analysis from the same sample demonstrated that sternal end phase morphology was at least twice as accurate.

Microscopic techniques have a number of disadvantages. Technically, this is a very laborious and time-consuming procedure requiring specialized training. Osteon counts are subject to a great deal of interobserver error. Even among experts, assessment of the osteons and related structures varies considerably and is thus subject to high levels of interobserver inconsistency. This technique also requires destruction of the long bone shaft or rib which must be cut to prepare thin sections. Therefore, this approach must be pursued with caution when dealing with forensic cases. It is not routinely used and is usually limited to specialists.

Discussion

Estimating age at death continues to be one of the most challenging aspects of skeletal analysis because of the complexity and individual variation of the aging process and the myriad of factors that affect it. Osteologists use the term 'estimate' because we can get close to chronological age – as with the rib phases which provide a meaningful range in which there is a high probability (95%) that true age will fall. It cannot be overemphasized that age estimates must be expressed as a range rather that erroneously using the mean value as a point estimate. To do so can lead to the elimination of the actual victim from being considered as a match for unknown remains.

Although scientists have gained a good idea about the general stages of age changes in the skeletal system that hold true for the most part, there are numerous factors that accelerate or decelerate the process. This can result in unexpected morphologic manifestations that may cause errors in age estimation. Human variation, on both the individual and population level is significant. So race and sex differences must always be considered in age estimation standards and decisions.

When estimating age from the skeleton, one must remember that what has been analyzed in research is the skeletal age rather than birth certificate-based chronological age. Researchers attempt to reconcile the internally regulated developmental and remodeling processes in the skeleton with the passage of time. Conflicts (errors in age estimation) arise because the former follows its own 'biological clock', and the latter is fixed by the regular, unalterable ticking of the atomic clock. To make matters worse, growth, development and remodeling can proceed at different rates in different parts of the skeleton. In other words, one component may not clearly reflect the sum total of skeletal development. Chronological age represents time based or true age and attempts to elucidate this depend on expected bone formations at a given point in time. However, a person at a particular level of skeletal development may have reached that stage earlier or later than the average. The physiologic processes that underlie aging are dependent on many internal and external influences including genetic make up, health and nutritional status, and substance abuse. During growth the situation becomes especially complex when age estimation is linked to bone lengths in children because no matter how much variation is incorporated, the potential is there to make errors. The same applies to remodeling in adulthood. This is why it is so important to provide age ranges so that extremes are not omitted from the police list of missing victims.

Finally, it must also be kept in mind that some skeletal sites and methods are more accurate than others. However, forensic personnel must be cognizant of all possible options since one rarely has a complete skeleton or even the bone of choice.

See also: **Anthropology**: Overview; Skeletal Analysis; Sex Determination; Determination of Racial Affinity; Excavation and Retrieval of Forensic Remains; Bone Pathology and Antemortem Trauma in Forensic Cases. Overview and Meaning of ID.

Further Reading

Acsádi G and Nemeskéri J (1970) *History of Human Life Span and Mortality*. Budapest: Akadémiai Kiadó.

Bang G and Ramm E (1970) Determination of age in humans from root dentin transparency. *Acta Odontologica Scandinavica* **28**:3–35.

El-Nofely A and İşcan MY (1989) Assessment of age from the dentition in children. In: İşcan MY (ed.) *Age Markers in the Human Skeleton*, pp. 237–254. Springfield, IL: Charles C Thomas.

Gustafson G (1950) Age determinations on teeth. *Journal of American Dental Association* **41**:45–54.

İşcan MY (ed.) (1989) *Age Markers in the Human Skeleton*. Springfield, IL: Charles C Thomas.

İşcan MY and Kennedy KAR (eds) (1989) *Reconstruction of Life from the Skeleton*. New York: Alan R. Liss.

İşcan MY and Loth SR (1989) The osteological manifestations of age in the adult. In: İşcan MY and Kennedy KAR (eds) *Reconstruction of Life from the Skeleton*, pp. 23–40. New York: Wiley-Liss.

İşcan MY and Loth SR (1997) The scope of forensic anthropology. In: Eckert WG (ed.) *Introduction to Forensic Sciences*. pp. 343–369. Boca Raton, FL: CRC Press.

İşcan MY, Loth SR and Scheuerman EH (1992) Age assessment from the sternal end of the rib and pubic symphysis: a systematic comparison. *Anthropoologie (Prague)* **30**:41–44.

İşcan MY, Loth SR and Wright RK (1984) Age estimation from the rib by phase analysis: White males. *Journal of Forensic Sciences* **29**:1094–1104.

İşcan MY, Loth SR and Wright RK (1985) Age estimation from the rib by phase analysis: White females. *Journal of Forensic Sciences* **30**:853–863.

İşcan MY, Loth SR and Wright RK (1987) Racial variation in the sternal extremity of the rib and its effect on age determination. *Journal of Forensic Sciences* **32**:452–466.

Johnston FE and Zimmer LO (1989) Assessment of growth and age in the immature skeleton. In: İşcan MY and Kennedy KA (eds) *Reconstruction of Life from the Skeleton*, pp. 11–22. New York: Alan R. Liss.

Klepinger LL, Katz D, Micozzi MS and Carroll L (1992) Evaluation of cast methods for estimating age from the os pubis. *Journal of Forensic Sciences* **37**:763–770.

Kósa F (1989) Age estimation from the fetal skeleton. In: İşcan MY (ed) *Age Markers in the Human Skeleton*, pp. 21–54. Springfield, IL: Charles C Thomas.

Krogman WM and İşcan MY (1986) *The Human Skeleton in Forensic Medicine*. Springfield, IL: Charles C Thomas.

Loth SR (1990) A comparative analysis of the ribs of Terry collection blacks. *Adli Tıp* Dergisi (Journal of Forensic Medicine), Istanbul **6**:119–127.

Loth SR and İşcan MY (1989) Morphological assessment of age in the adult: the thoracic region. In: İşcan MY (ed.) *Age Markers in the Human Skeleton*, pp. 105–135. Springfield: IL: Charles C Thomas.

Loth SR and İşcan MY (1994) Morphological indicators of skeletal aging: Implications for paleodemography and paleogerontology. In: Crews D and Garruto R (eds) *Biological Anthropology and Aging: An Emerging Synthesis*, pp. 394–425. New York: Oxford University Press.

Loth SR, İşcan MY and Scheuerman EH (1994) Intercostal variation at the sternal end of the rib using the present standards. *Forensic Science International* **65**:135–143.

Lovejoy CO (1985) Dental wear in the Libben population: its functional pattern and role in the determination of adult skeletal age at death. *American Journal of Physical Anthropology* **68**:47–56.

Massler M, Schour I and Poncher HG (1941) Developmental pattern of the child as reflected in the calcification pattern of teeth. *American Journal of Diseases of Children* **62**:33.

Meindl RS and Lovejoy CO (1985) Ectocranial suture closure: a revised method for the determination of skeletal age at death based on the lateral-anterior sutures. *American Journal of Physical Anthropology* **68**:57–66.

Moorrees CFA, Fanning EA and Hunt Jr. EE (1963) Age variation of formation stages in ten permanent teeth. *Journal of Dental Research* **42**:1490–1502.

Murray KA and Murray T (1991) A test of the auricular surface aging technique. *Journal of Forensic Sciences* **36**:1162–1169.

Novotný V, İşcan MY and Loth SR (1993) Morphologic and osteometric assessment of age, sex, and race from the skull. In: İşcan MY and Helmer RP (eds) *Forensic Analysis of the Skull: Craniofacial Analysis, Reconstruction, and Identification*, pp. 71–88. New York: Wiley-Liss.

Russell KF, Simpson SW, Genovese J, Kinkel MD, Meindl RS and Lovejoy CO (1993) Independent test of the fourth rib aging technique. *American Journal of Physical Anthropology* **92**:53–62.

Sex Determination

S R Loth, Department of Anatomy, University of Pretoria, Pretoria, South Africia
M Y İşcan, Adli Tıp Enstitusu, Instanbul Univesitesi, Istanbul, Turkey

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0756

Introduction

Assessment of sex is one of the most vital determinations to make when it is necessary to establish identity from skeletal remains. In the first place, this classification effectively cuts the number of possible matches in half, and proper identification could never be made if officials were told to look for a male when the remains were actually those of a female. Moreover, identification techniques such as facial reconstruction would be impossible if sex could not be correctly established. Thus, isolating, interpreting and quantifying the manifestations of sex is an essential part of all skeletal analyses. Unfortunately, this is often not a simple process since male and female attributes span a continuum of morphologic configurations and metric measures in the skeleton. Although some bones are

İşcan MY (ed.) (1989) *Age Markers in the Human Skeleton*. Springfield, IL: Charles C Thomas.

İşcan MY and Kennedy KAR (eds) (1989) *Reconstruction of Life from the Skeleton*. New York: Alan R. Liss.

İşcan MY and Loth SR (1989) The osteological manifestations of age in the adult. In: İşcan MY and Kennedy KAR (eds) *Reconstruction of Life from the Skeleton*, pp. 23–40. New York: Wiley-Liss.

İşcan MY and Loth SR (1997) The scope of forensic anthropology. In: Eckert WG (ed.) *Introduction to Forensic Sciences*. pp. 343–369. Boca Raton, FL: CRC Press.

İşcan MY, Loth SR and Scheuerman EH (1992) Age assessment from the sternal end of the rib and pubic symphysis: a systematic comparison. *Anthropoologie (Prague)* **30**:41–44.

İşcan MY, Loth SR and Wright RK (1984) Age estimation from the rib by phase analysis: White males. *Journal of Forensic Sciences* **29**:1094–1104.

İşcan MY, Loth SR and Wright RK (1985) Age estimation from the rib by phase analysis: White females. *Journal of Forensic Sciences* **30**:853–863.

İşcan MY, Loth SR and Wright RK (1987) Racial variation in the sternal extremity of the rib and its effect on age determination. *Journal of Forensic Sciences* **32**:452–466.

Johnston FE and Zimmer LO (1989) Assessment of growth and age in the immature skeleton. In: İşcan MY and Kennedy KA (eds) *Reconstruction of Life from the Skeleton*, pp. 11–22. New York: Alan R. Liss.

Klepinger LL, Katz D, Micozzi MS and Carroll L (1992) Evaluation of cast methods for estimating age from the os pubis. *Journal of Forensic Sciences* **37**:763–770.

Kósa F (1989) Age estimation from the fetal skeleton. In: İşcan MY (ed) *Age Markers in the Human Skeleton*, pp. 21–54. Springfield, IL: Charles C Thomas.

Krogman WM and İşcan MY (1986) *The Human Skeleton in Forensic Medicine*. Springfield, IL: Charles C Thomas.

Loth SR (1990) A comparative analysis of the ribs of Terry collection blacks. *Adli Tıp* Dergisi (Journal of Forensic Medicine), Istanbul **6**:119–127.

Loth SR and İşcan MY (1989) Morphological assessment of age in the adult: the thoracic region. In: İşcan MY (ed.) *Age Markers in the Human Skeleton*, pp. 105–135. Springfield: IL: Charles C Thomas.

Loth SR and İşcan MY (1994) Morphological indicators of skeletal aging: Implications for paleodemography and paleogerontology. In: Crews D and Garruto R (eds) *Biological Anthropology and Aging: An Emerging Synthesis*, pp. 394–425. New York: Oxford University Press.

Loth SR, İşcan MY and Scheuerman EH (1994) Intercostal variation at the sternal end of the rib using the present standards. *Forensic Science International* **65**:135–143.

Lovejoy CO (1985) Dental wear in the Libben population: its functional pattern and role in the determination of adult skeletal age at death. *American Journal of Physical Anthropology* **68**:47–56.

Massler M, Schour I and Poncher HG (1941) Developmental pattern of the child as reflected in the calcification pattern of teeth. *American Journal of Diseases of Children* **62**:33.

Meindl RS and Lovejoy CO (1985) Ectocranial suture closure: a revised method for the determination of skeletal age at death based on the lateral-anterior sutures. *American Journal of Physical Anthropology* **68**:57–66.

Moorrees CFA, Fanning EA and Hunt Jr. EE (1963) Age variation of formation stages in ten permanent teeth. *Journal of Dental Research* **42**:1490–1502.

Murray KA and Murray T (1991) A test of the auricular surface aging technique. *Journal of Forensic Sciences* **36**:1162–1169.

Novotný V, İşcan MY and Loth SR (1993) Morphologic and osteometric assessment of age, sex, and race from the skull. In: İşcan MY and Helmer RP (eds) *Forensic Analysis of the Skull: Craniofacial Analysis, Reconstruction, and Identification*, pp. 71–88. New York: Wiley-Liss.

Russell KF, Simpson SW, Genovese J, Kinkel MD, Meindl RS and Lovejoy CO (1993) Independent test of the fourth rib aging technique. *American Journal of Physical Anthropology* **92**:53–62.

Sex Determination

S R Loth, Department of Anatomy, University of Pretoria, Pretoria, South Africia
M Y İşcan, Adli Tıp Enstitusu, Instanbul Univesitesi, Istanbul, Turkey

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0756

Introduction

Assessment of sex is one of the most vital determinations to make when it is necessary to establish identity from skeletal remains. In the first place, this classification effectively cuts the number of possible matches in half, and proper identification could never be made if officials were told to look for a male when the remains were actually those of a female. Moreover, identification techniques such as facial reconstruction would be impossible if sex could not be correctly established. Thus, isolating, interpreting and quantifying the manifestations of sex is an essential part of all skeletal analyses. Unfortunately, this is often not a simple process since male and female attributes span a continuum of morphologic configurations and metric measures in the skeleton. Although some bones are

better indicators than others, there is no skeletal feature that rivals the definitiveness of differences between fleshed individuals.

In some cases, remains of clothing, hair, jewelry and other artifacts may be helpful to corroborate a skeletal assessment and indeed lead to a positive identification of the items in question, but they may also be misleading. Today, long hair is not always a female prerogative, nor is a man's shirt only worn by males. A woman may be wearing her boyfriend's ring and both sexes often wear earrings. Cross-dressers are also a problem and identification may be delayed or even impossible without a thorough skeletal examination. One must be particularly careful in mass disasters where both bones and personal items may be comingled.

Sexual dimorphism

In normal, living humans, sex is a discrete trait determined by the actions of the XX or XY genotype and diagnosed by observing one of only two possible morphological features. These differences (e.g. external genitalia) are apparent at birth and clearly recognizable even during the prenatal period. In the skeleton, however, the most effective sex indicators do not begin to develop until adolescence, and some are not fully expressed until adulthood. Although almost every human bone has been analyzed, both metrically and morphologically, even a complete, mature pelvis cannot consistently guarantee more than 95% separation of the sexes.

There are two methodological approaches to sexing skeletal remains: morphological and osteometric. Morphologic techniques focus on shape – the bony configurations that are macroscopically visible and differ between males and females. There are important advantages to this approach, especially when a particular form is recognizable despite temporal and population variation. Obvious morphological differences such as the subpubic angle of the pelvis and ramus flexure in the mandible allow optimal separation of the sexes (approaching 95% accuracy). One reason these are so effective is that they are developmental in nature. Subpubic angle expansion is one of several pelvic features that develop during adolescence to accommodate the demands of childbearing on the female pelvis. Ramus flexure results from postadolescent growth in males, whereas the lack of it maintains the straight juvenile form in females. The main drawback of the morphologic approach in general (especially when judgement of size is involved) is that it is based on 'eye balling' so if the formation in question is not patently obvious, experience becomes an essential component. The observer must develop a sense of what is relatively large or small, angled or curved, wide or narrow.

Unfortunately, not all skeletal components have consistent and discernible morphologic evidence of sexual dimorphism. Problems arise when differences are only size based and there is no clear male or female shape. Many bony sites are known to exhibit visible differences in size, but osteologists can only determine if they fall into the male or female range after years of experience. However, the overlap between the sexes, along with significant population variation, makes this problematic for all but extreme cases. Although the size of the mastoid process, for example, is often used for sexing, it is difficult to judge by simply looking at it, even if one is familiar with the range of variability within a given group, e.g. Southern African Blacks have much smaller mastoids than Whites or American Blacks. Therefore, metric analysis, based on bone dimensions, is the method of choice for skeletal parts like long bones that do not exhibit clearly definable shape variants. Often a combination of measurements are selected from each bone to maximize sex diagnosis using discriminant function statistics. The major problem with this technique is that standards are temporally sensitive and population specific (both between and within groups) – a formula for Whites may not work on Blacks or Asian Whites, and formulas for American Blacks may not be effective on South African Blacks. Another drawback is the metric overlap of the sexes which can be as high as 85%. Finally, although the metric approach draws heavily on the fact that males tend to be more robust because normal testosterone levels produce greater muscle mass, cultural variation in functional demands may lead to both differences in the expression of sexual dimorphism and its magnitude. This is of particular concern in bones where dimorphism is not primarily developmental, i.e. those that reflect weight bearing and labor-intensive stress.

The diagnosis of sex from the skeleton is further complicated by a number of diverse factors ranging from environmental influences and pathologic conditions to temporal change, diet and occupational stress. In addition, recent genetic and endocrine studies have demonstrated that the expression of the sex genotype can be modified by a malfunction of the appropriate hormonal triggers or on-site receptors. As noted before, all skeletal sites and traits are not equally effective at distinguishing one sex from the other. So the question arises – why bother with the less definitive parts? The answer is simple – skeletal remains are rarely complete and undamaged, even in recent forensic cases. Therefore, the investigator must be able to glean as much information as possible from any bone or fragment. This article presents an over-

view of current methods and their accuracy for sex determination from the human skeleton.

The Immature Skeleton

Because sex differences in the immature skeleton are not readily observable before puberty, the few attempts that have been made to look for dimorphic indicators have met with limited success, especially for individuals in the first few years of life. One exception is a recent study of the development of sexual dimorphism in the mandible that revealed differences in shape of the inferior symphyseal border and corpus beginning with the eruption of the central incisors at about seven months until about 4–5 years. As can be seen at the top of **Fig. 1**, some mandibles had a gradually curved or rounded border – these turned out to be female. In contrast, the anterior aspect of the male mandibles (below) extends abruptly downward coming to a point or squaring off at the symphysis. These mandibles clearly demonstrate the sharply angled transition to the lateral body in males so different from the gradual transition in the female. Mandibular corpus (body) shape (not to be confused with dental arcade form) is another important feature. The female is characterized by a distinctly rounded contour of the body resulting from a gradual transition from front to sides, in contrast to the male in whom the sides diverge in a sharp, angular fashion from a nearly horizontal anterior region just past the canines. Even when the male symphyseal base is more rounded, there is still a precipitous descent from the lateral body. Accuracy which averages 81% is higher than that attainable from adult chin shapes or any other part of the immature skeleton in this age range. By 6–8 years of age, traditional adult chin morphology (squared or undercut in males; rounded in females) was recognizable. The anterior mandible was found to be nearly identical to adults in both form and size by as early as 13. The 69% predictive accuracy in 6–19 year olds was not significantly different from the 71% obtained in adults aged 20 and older.

Attempts to capitalize on sex differences in the size of both the deciduous and adult dentition have not been successful. Recent work has demonstrated that tooth dimensions (length and breadth) do not yield a statistically significant separation of the sexes. Another study attempted to determine sex in the fetal, newborn and 6-month-old ilium. Based on observations that the auricular surface was elevated in males accuracy was reported to be 75% in females and 92% in males in this age group. Metrically, however, no sex difference was found in various dimensions of the ilium in this age group. Sexual differentiation of the pelvis is not clearly defined until adolescence. As detailed below, expected variation in the shape of the pubic bone, sciatic notch and subpubic angle are manifest by the mid-teens in normally developing individuals.

Figure 1 Sex differences in the immature mandible. Female morphology (above) has a rounded corpus shape with a gradual transition from the lateral body to the symphysis. Males (below) show a steep abrupt transition with an angular corpus (not dental arcade) shape

The Adult Skeleton: Morphology

When the morphologic expressions of sexual dimorphism in the adult skeleton are easily distinguishable they provide the most consistently reliable means of diagnosing sex. The best places to look for these formations are the skull and pelvis.

The skull

Sex differences have been investigated in nearly every feature of the skull. The most dimorphic part of the skull is the mandible. **Fig. 2** illustrates sex-linked morphologic configurations of the mandibular ramus. In adult males, ramus shape is characterized by a distinct rearward angulation of the posterior border of the ramus at the level of the occlusal surface of the molars (Fig. 2); this is referred to as 'ramus flexure'. Females, however, are not flexed at this part of the bone – the posterior border maintains the straight juvenile shape, or, in some females, angulation occurs at a higher point at the neck of the condyle (Fig. 2). This single trait has demonstrated diagnostic accuracy of 91–99% with confirmation by independent tests.

If only the symphyseal region of the mandible is present, a squared or undercut base is often associated with males and a round or pointed inferior outline with females. In modern Whites and Blacks there is an overlap of the female shapes in the males and accuracy is only about 70%. Research has demonstrated that adult chin shapes are highly heritable and only coincidentally linked to sex. The same appears to be true for gonial eversion, except that this condition does not even attain separation at levels greater than chance (49%). A better assessment can be made from mandibular condyle shape (68%). Large, thick, oblong condyles are associated with males, and small, round or thin, oblong shapes are usually female.

Table 1 contains a list of dimorphic cranial characteristics (**Fig. 3**). In general, the male skull is expected to have more rugged sites of muscle attachments, a prominent glabellar region, and thicker bones. A large, pronounced supraorbital torus and rugged occipital protuberance can be considered reliable male characteristics. When viewed laterally, the sloping male forehead gives the vault a more rounded outline than that of the relatively vertical female forehead. Males have larger mastoid processes. Both frontal and parietal eminences tend to be better developed in females. Some differences in the cranium are of a proportional nature such as the orbits. Because eyes are of similar size in both sexes, the orbits appear

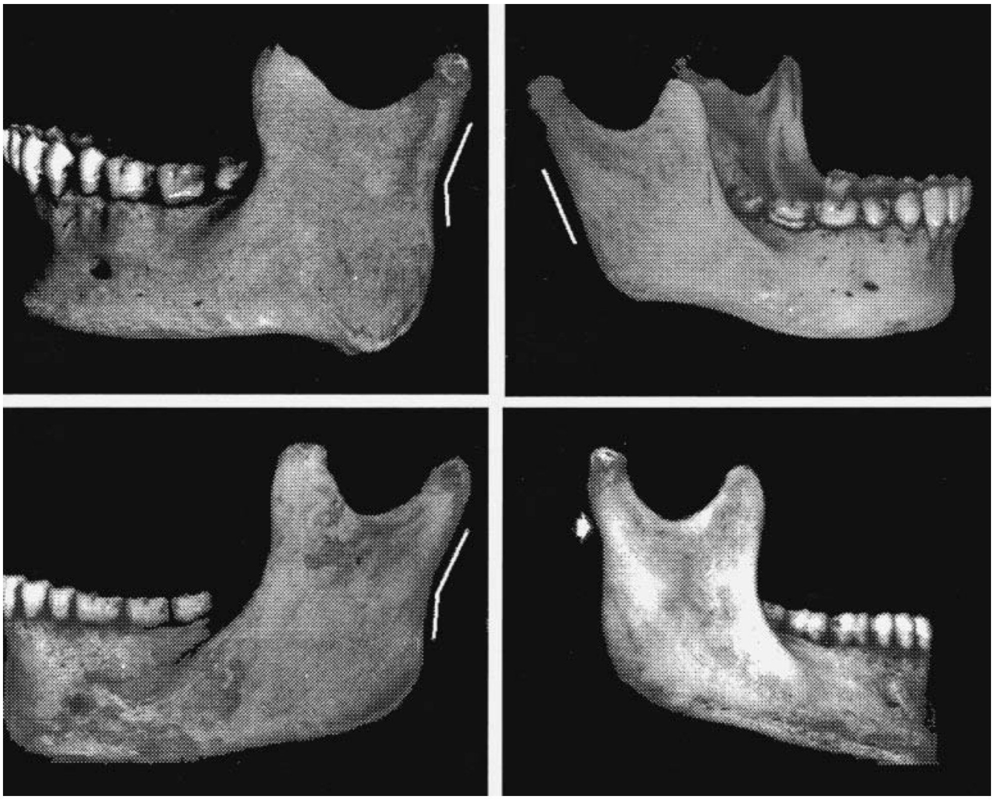


Figure 2 Sex difference in the adult mandible. The posterior ramus of adult males (left) is flexed at the level of the occlusal surface of the molars. The female ramus maintains the straight juvenile shape (top right) or may flex above the occlusal level near the neck of the condyle (lower right).

Table 1 Morphologic sex differences in skull and pelvis

Trait	*Male*	*Female*
General appearance	Rugged	Smooth
Supraorbital ridges	Medium to large	Small to absent
Mastoid processes	Medium to large	Small to medium
Occipital area	Muscle lines and protuberance marked or hooked	Muscle lines and protuberance not marked
Frontal eminences	Small	Large
Parietal eminences	Small	Large
Orbits	Rectangular	Rounded
	Small relative to face	Large relative to face
	Rounded margins	Sharp margins
Forehead	Sloped	Vertical
Mandibular ramus flexure	Ramus flexure	Straight ramus
Palate	Larger, broader, tends to U-shape	Small, tends to parabola
Teeth	Large, lower M1 more often 5 cusped	Small, molars often 4 cusped
Pubic symphysis	Higher	Lower
Pubic bone	Triangular	Square or rectangle
Subpubic angle	Narrow V-shaped, acute angle	Wide U-shaped, obtuse angle
Acetabulum	Large, tends to be directed laterally	Small, tends to be directed anterolaterally
Greater sciatic notch	Narrow, deep	Wide, shallow
Ischiopubic rami	Rough everted margin	Gracile, narrows near symphysis
Sacroiliac joint	Large	Small, oblique
Postauricular space	Narrow	Wide
Preauricular sulcus	Rarely present	Often present, well developed
Iliac tuberosity	Large, not pointed	Small or absent, pointed or varied
Sacrum	Longer, narrower with more evenly distributed curvature; often 5 or more segments	Shorter, broader, with tendency to marked curvature at S1–2 S2–5; 5 segments the rule
Pelvic brim or inlet	Heart-shaped	Circular, elliptical

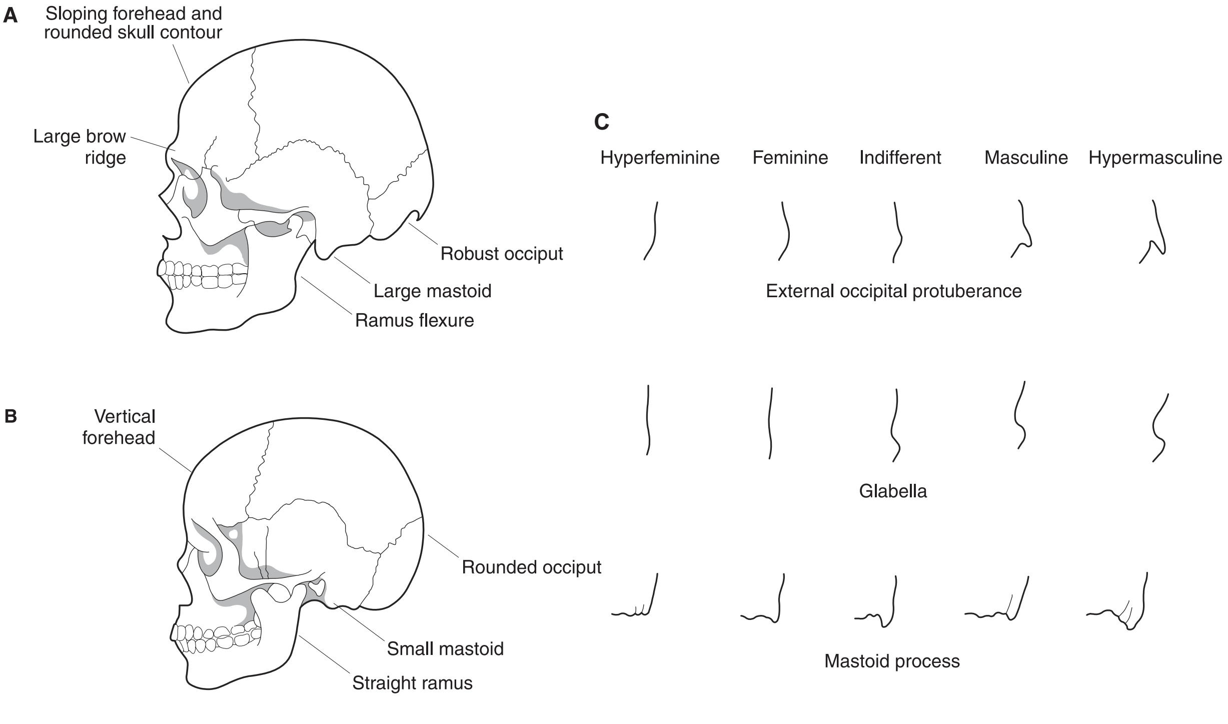


Figure 3 Sex differences in the skull (see Table 1). (**A**) Male features; (**B**) female features; (**C**) morphologic gradations.

to take up a larger portion of the face in females. Females also tend to have round or square orbits with sharp superior orbital margins, in contrast to the more rectangular orbits with rounded margins in males.

In all skeletal assessments, it is important to remember that humans vary both temporally and spatially. To this one may add variation arising from growth factors, environmental stress and pathology or trauma. Therefore, small or absent brow ridges or prominent eminences are not necessarily female indicators. The same can be said of a rounded chin and a less rugged occipital region, both of which are not uncommonly seen in males. In this regard, population differences are a major cause of variability in the expression of sexual dimorphism.

The pelvis

The pelvic girdle is composed of three bones: two coxal (or hip) bones and the sacrum (**Fig. 4**). Each hip bone is in turn the product of the fusion of three bones: the ilium, ischium and pubis. As mentioned earlier, the female pelvis must be designed to accommodate childbirth whereas the only requirements for males are support and locomotion. These differences are manifest in various parts of this complex structure. Table 1 summarizes the morphologic differences in the pelvis.

The most easily assessable morphologic feature is the subpubic angle formed by the ventral articulation of the pubic bones at the symphysis (Fig. 4). The male is characterized by roughly triangular pubic bones that form a narrow subpubic angle. Pubertal growth modifications in females are centered on the facilitation of childbirth. Elongation of the pubic bones creates a more rectangular shape and results in a much wider subpubic angle and larger outlet. Further enlargement of the pelvic outlet is accomplished by widening of the greater sciatic notch in females, but this feature may show considerable variability and is sometimes difficult to assess. If the sacrum can be articulated, it gives a much clearer picture of the width of the notch by completing the dorsal portion. Other female adaptations which can be used in sex determination include a rounded (as opposed to the heart-shaped male) inlet, a dorsally oriented tilt of the sacrum, a small, and everted ischial spine. The upper section or 'false pelvis' is broader and deeper in males whereas the lower portion or 'true pelvis' is wider in females. The ischial tuberosity is positioned more dorsally in males and laterally in females. A uniquely male trait is the rough everted area of the medial border of the ischiopubic ramus for the attachment of the crus (corpus cavernosum) of the penis. This ramus is more gracile in females and narrows as

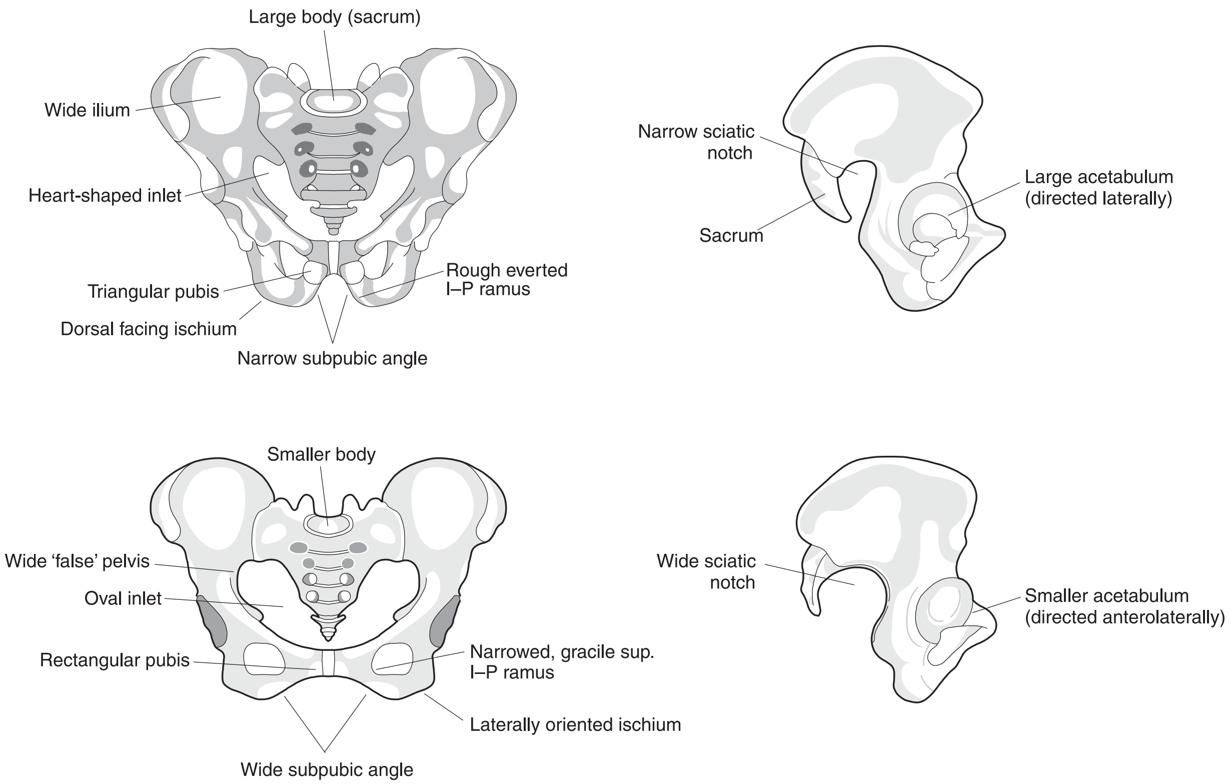


Figure 4 Sex differences in the pelvis (see Table 2). Male above, female below.

it approaches the symphysis. In the sacrum, females show a proportionally smaller superior articular surface relative to the total width of the wings.

As in all skeletal remains, the most diagnostic parts are not always available, especially following extended periods of exposure to the elements or years of burial. Since spongy bones are more prone to destruction, highly dimorphic bones like the pubis may not be available for diagnosis. Other clues to sex can be found in the sacroiliac joint and relate to its greater mobility in females. In many individuals there are accessory articular facets in males on both the iliac tuberosity and dorsal sacrum. If the sacrum is fused to the ilium, the individual was most likely male, in the absence of trauma or disease.

Other postcranial indicators

If the most reliable sites are missing, at least some indication of sex can also be found in certain features of the long bones and ribs. A foramen in the olecranon fossa of the humerus has been linked to females. If a long bone is large and robust with pronounced muscle attachments these male indicators take precedence. Rib morphology also shows sex differences. If the sternal end has large projections at the superior/inferior borders, this is a male characteristic. Postmenopausal females may also exhibit this phenomenon, but the projections are much thinner. It must be kept in mind that these fragile projections are often lost or broken. If the costochondral junction has bony projections extending from the floor of the pit, the individual is likely to be female.

The Adult Skeleton: Metric Analysis

Osteometric analysis can also yield high levels of accuracy for the diagnosis of sex. Techniques range from the calculation of a simple sectioning point derived from a single measurement to complex multivariate discriminant function analysis. Indices formed by the relationship of one dimension to another like that of the ischiopubic index allow male/female comparisons while eliminating size as a factor. This is important since the sexes can have as much as an 85% overlap for size alone.

Discriminant function analysis is one of the most commonly used techniques to develop sex determination formulas using one or more measurement from the skeleton. In general selection of dimensions for a formula depends on levels of intercorrelation as well as the degree of difference between the sexes. It is, for example, very likely that femoral distal breadth is significantly correlated with tibial proximal breadth and therefore one of these may suffice to provide the best result. In long bones it has been observed that epiphyseal measurements are better indicators of sex than length or diaphyseal dimensions. It has also been shown that, metrically, sexual dimorphism in the long bones is more diagnostic than in the skull and necessitates the use of fewer dimensions. For example, one may need several skull measurements to obtain a sexing accuracy as high as that of the femoral or humeral head diameter alone.

With this in mind, univariate discriminant functions were calculated using eight single dimensions from the humerus, femur and tibia (**Table 2**). Although most of the dimensions used are standard measurements (clearly defined in major reference books) these dimensions were also selected because these segments of the skeleton are commonly present at a crime scene or grave even if the skeleton is badly fragmented. In order to make this methodology applicable to diverse populations, osteometric analyses of modern (twentieth century) skeletal samples have been conducted on US Whites and Blacks, South African Whites and Blacks, and Asians including mainland Chinese and Japanese. Table 2 lists descriptive statistics for males and females along with the average of the two sexes (sectioning point) and percentage accuracy. In all samples, males were significantly larger than females. Using this technique, determination of sex is accomplished by comparing the dimension of an unknown bone with the sectioning point for that population. For example, if the humeral head diameter of an unknown American White is 47 mm, classification would be 'male' (any value larger than 45.6 mm classifies a person as male in this group). Although overall accuracy is 84.8% (Table 2), the farther the measurement is from the sectioning point, the greater the likelihood of correct sexing. Moreover, if the humeral head diameter is greater than the male mean (48.6 mm) or less than the female mean (42.6 mm), the probability of correct determination would be much higher. It must be emphasized however, that it is necessary to know racial affinity in these cases unless the values of the unknown specimen are extreme.

As noted earlier, metrics are group specific. As can be seen in these figures, populations differ from each other in the degree and range of dimorphism exhibited by various dimensions. This reflects differences in both size and proportions. In general, South African Whites are the largest of the groups assessed whereas the average South African Black is among the smallest with a mean closer to that of the Chinese and Japanese. Table 2 also reveals population differences in the locus of dimorphism. For example, femoral distal breadth is the most diagnostic measurement in South African Whites, whereas maximum dimorphism in

Table 2 Metric determination of sex and classification accuracy from single long bone dimensions (mm) in American Whites and Blacks, South African Whites and Blacks, and Chinese and Japanese.

	Terry Whites		*Terry Blacks*		*S Afr Whites*		*S Afr Blacks*		*Chinese*		*Japanese*	
Dimensions	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
n	Male: 46 Female: 45		Male: 45 Female: 44		Male: 44 Female: 42		Male: 37 Female: 44		Male: 33 Female: 29		Male: 39 Female: 32	
Humeral head diameter												
Males	48.6	4.18	48.0	3.17	49.0	3.28	43.8	2.23	45.1	2.56	44.0	1.72
Females	42.4	2.33	42.8	2.29	43.0	2.28	37.7	2.08	39.3	2.20	39.1	2.67
Average	45.6	4.60	45.4	3.81	46.1	4.13	40.5	3.73	42.4	3.76	41.8	3.27
% Accuracy	84.8		86.8		84.6		90.9		80.5		87.3	
Humeral epicondylar breadth												
Males	63.9	4.35	64.5	4.42	63.6	3.67	61.3	6.43	58.5	3.87	59.9	2.11
Females	56.3	3.52	57.5	3.02	55.7	2.69	53.5	3.62	52.2	2.75	52.2	3.95
Average	60.2	5.50	61.0	5.13	59.7	5.10	57.0	10.05	55.6	4.62	56.4	4.91
% Accuracy	83.7		86.8		88.4		88.6		84.4		91.1	
Femoral head diameter												
Males	48.9	3.14	48.1	3.23	48.4	2.76	45.3	2.33	46.2	2.60	46.0	1.86
Females	42.8	2.39	42.6	2.03	43.1	2.13	39.2	2.60	41.0	2.31	40.8	2.54
Average	45.9	4.13	45.4	3.86	45.8	3.64	42.0	3.94	43.8	3.58	43.6	3.40
% Accuracy	82.6		88.0		85.9		91.9		83.1		88.6	
Femoral midshft circumference												
Males	91.5	4.82	92.7	5.65	92.0	5.47	90.0	5.16	85.3	6.37	85.4	4.16
Females	82.8	5.67	84.0	5.73	84.5	5.14	78.8	4.74	75.3	4.66	78.3	6.40
Average	87.2	6.80	88.4	7.16	88.3	6.48	84.0	7.45	80.6	7.53	82.2	6.32
% Accuracy	79.1		79.1		74.3		89.9		81.7		78.5	
Femoral distal breadth												
Males	83.1	4.44	82.8	4.68	84.0	4.61	79.2	4.24	80.2	4.47	81.1	2.73
Females	75.3	3.58	74.7	3.66	75.2	3.27	69.6	4.78	69.8	3.00	72.4	4.31
Average	79.2	5.62	78.8	5.85	79.7	5.98	74.0	6.60	75.3	6.50	77.1	5.58
% Accuracy	79.4		87.0		90.5		86.4		92.3		85.5	
Tibial proximal breadth												
Males	76.0	3.68	77.1	4.14	78.2	4.43	74.9	3.72	72.8	3.74	73.5	2.87
Females	68.6	3.69	68.3	3.00	69.8	2.96	64.9	7.37	63.6	3.36	66.2	4.42
Average	72.3	5.25	72.8	5.68	74.1	5.63	69.5	7.80	68.5	5.81	70.2	5.16
% Accuracy	85.9		89.1		87.7		86.5		89.5		88.6	
Tibial circumference at nutrient foramen												
Males	96.2	5.76	100.1	6.91	97.3	6.13	98.4	6.35	93.1	6.52	91.6	4.90
Females	86.7	7.90	90.0	6.20	87.2	6.13	85.1	5.20	76.4	5.77	82.5	8.56
Average	91.5	8.36	95.1	8.28	92.4	7.94	91.2	8.78	85.3	10.41	87.5	8.13
% Accuracy	76.1		80.4		82.1		83.50		90.4		80.0	
Tibial distal breadth												
Males	47.7	3.15	47.6	3.76	46.8	2.62	45.4	2.56	45.1	2.38	45.3	2.18
Females	43.3	2.86	47.6	2.70	41.9	2.50	39.9	2.23	39.0	2.34	40.7	2.41
Average	45.5	3.71	45.4	3.98	44.4	3.53	42.4	3.63	42.2	3.87	43.3	3.24
% Accuracy	78.3		80.4		80.7		87.5		90.1		82.5	

the Japanese is found in the epicondylar breadth of the humerus.

Parturition Scars

In the last 30 years, there has been considerable debate as to whether one can determine if pregnancy and parturition leave traces in the skeleton. The initial hypothesis was that these processes can tear the ligaments of the pubic bones and sacroiliac joint and are responsible for causing scars or pits at these sites. These manifestations are found on the dorsal surface of the pubis and preauricular sulcus at the inferior margin of the auricular surface of the ilium. It was also claimed that the number of births can be estimated by the size and number of so-called ‘parturition pits’. However, more recent studies using bones from females of known parity disagreed with this hypothesis. The scars were found in many females who had never had children, and scar size and shape showed variation; these, too, were not related to the number of children borne by the

woman. Obviously, other factors must cause similar scarring. These may include horseback riding, habitual squatting, and exercise involving hopping and jumping.

Summary and Conclusions

Although there are many techniques to determine sex from practically every part of the skeleton, there are a number of caveats. The accuracy of both morphological and osteometric sex assessment depends on rigorous training and first-hand experience in observing skeletal remains. The overlap between the sexes implies that structures show gradations of differences within and between sexes and populations. This gradation must be carefully evaluated by studying large documented skeletal collections from different regions of the world. Bone measurement is a tedious process and requires consistency and standardization. The technique has a long history and although there is international agreement among scientists on how measurements are taken (to allow comparability), individual variation in measuring and interobserver error are always troublesome factors.

Even though decades of research have resulted in a plethora of methods and standards of assessment, many factors complicate the determination of sex from the skeleton. Age must be considered because the skeleton does not exhibit maximum dimorphism until sexual maturation and growth are complete. Problems also arise when one only has a partial or fragmented skeleton. It is not always easy to distinguish certain bones of a male of 17 from one in his 20s and since growth can continue into the latter decade, sex differences may neither be clear metrically nor morphologically. This is further complicated by the fact that there is significant variation between populations within each race who inhabit different geographic regions and constitute established gene pools.

Finally, since anything that affects the growth, development, maintenance and remodeling of the skeleton can modify the expression of sexual dimorphism, trauma and disease can result in missexing. Some conditions cause abnormal bony enlargement whereas others result in atrophy or obliterate normally clear morphologic differences. The intimate relationship between bone and muscle means that even in the absence of obvious damage to a skeletal component, muscular impairment can lead to inappropriate remodeling. All of these factors, above and beyond the normally occurring variability, subtleness, and overlap of features in the human skeleton, make it essential that only properly trained and experienced forensic anthropologists be consulted in every case involving the medicolegal system.

See also: **Anthropology**: Overview; Skeletal Analysis; Morphological Age Estimation; Determination of Racial Affinity. **Identification/Individualization**: Overview and Meaning of ID.

Further Reading

Acsádi G and Nemeskéri J (1970) *History of Human Life Span and Mortality*. Budapest: Akadémiai Kiadó.

Bennett KA (1981) On the expression of sex dimorphism. *American Journal of Physical Anthropology* **56**:59–61.

Galloway A (1995) Determination of parity from the maternal skeleton: an appraisal. *Rivista di Antropologia* **73**:83–98.

Introna FJ, Cantatore F, Dragone M and Colonna M (1993) Dimorfismo sessuale dei denti decidui nell'identificazione medico-legale. *Bolletino della Societ'a Italiana di Biologia Sperimentale* **69**:223–230.

İşcan MY and Loth SR (1997) The scope of forensic anthropology. In: Eckert WG (ed.) *Introduction to Forensic Sciences*, pp. 343–369. Boca Raton: CRC Press.

İşcan MY, Yoshino M and Kato S (1995) Sexual dimorphism in modern Japanese crania. *American Journal of Human Biology* **7**:459–464.

İşcan MY, Loth SR, King CA, Ding S and Yoshino M (1998) Sexual dimorphism in the humerus: a comparative analysis of Chinese, Japanese and Thais. *Forensic Science International* **98**: (in press).

Krogman WM and İşcan MY (1986) *The Human Skeleton in Forensic Medicine*. Springfield: Charles C. Thomas.

Loth SR (1996) *Sexual dimorphism in the human mandible: a developmental and evolutionary perspective*. PhD Dissertation. Johannesburg, South Africa: University of the Witwatersrand.

Loth SR and Henneberg M (1996) Mandibular ramus flexure: a new morphologic indicator of sexual dimorphism in the human skeleton. *American Journal of Physical Anthropology* **99**:473–485.

Rösing FW (1983) Sexing immature skeletons. *Journal Human Evolution* **12**:149–155.

St Hoyme L and İşcan MY (1989) Determination of sex and race: accuracy and assumptions. In: İşcan MY and Kennedy KAR (eds) *Reconstruction of Life from the Skeleton*, pp. 53–94. New York: Wiley-Liss.

Stewart TD (1979) *Essentials of Forensic Anthropology: Especially as Developed in the United States*. Springfield: Charles C Thomas.

Steyn M and İşcan MY (1997) Sex determination from the femur and tibia in South African whites. *Forensic Science International* **90**:111–119.

Weaver DS (1980) Sex differences in the ilia of a known sex and age sample of fetal and infant skeletons. *American Journal of Physical Anthropology* **52**:191–195.

Workshop of European Anthropologists (1980) Recommendations for age and sex diagnoses of skeletons. *Journal of Human Evolution* **9**:517–549.

Skeletal Analysis

N J Sauer and **W L Lackey**,
Department of Anthropology, Michigan State University, East Lausing, MI, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0609

Introduction

Because they are so durable, bones and teeth are often pivotal to the solution of human remains cases. The skin, flesh, organs and body fluids (the domain of forensic pathologists) that provide such a wealth of data for identification and the interpretation of events that surround death, normally disappear after a few months or weeks. But the skeleton may endure, often intact, for centuries. Of course, in events like fires, remains may become skeletonized within minutes. What can be learned from a skeleton? How can skeletal remains contribute to forensic science?

Forensic anthropologists are specialists in human skeletal biology who apply their skills to medicolegal cases. Typically, forensic anthropologists have obtained an advanced degree (Masters or PhD) in physical anthropology with a specialization in skeletal biology and training in the legal applications of their skills. In North America most forensic anthropologists are members of the Physical Anthropology Section of the American Academy of Forensic Sciences. About 50 North American forensic anthropologists have been certified by (and are Diplomates of) the America Board of Forensic Anthropology. Board Certification requires a PhD in Physical Anthropology (human osteology specialization), a minimum of 3 years of practical experience, and the successful completion of written and practical exams.

Currently, most forensic anthropologists are employed by universities, however, some work for local, state or federal government agencies. The United States Central Identification Laboratory in Hawaii (CILHI), for example, has a staff that includes more than 20 full-time forensic anthropologists. The CILHI anthropologists help carry out the mission of identifying the remains of Americans found on foreign soil. A few forensic anthropologists assist the United Nations and the organization, Physicians for Human Rights, with the identification of suspected victims of political activity in places like Latin America, the former Yugoslavia and Central Africa.

Investigators who bring remains to forensic anthropologists usually want one or several of the following questions answered: Are these remains human? Are they recent enough in origin to be of medicolegal interest? Who do the remains represent? Is there evidence for cause or manner of death? This article illustrates the most important kinds of information that can be gained from studying the skeleton and illustrate how such information can contribute to forensic science.

Identification

Are the remains human?

Forensic anthropologists are experts in the morphology (form) of the human skeleton. Many suspected human cases are decided in the few minutes that it takes the anthropologist to recognize that a single bone or bone fragment does not come from a human skeleton. The ability to quickly and confidently make such a judgment results from years of intensive study of the form and function of human bone. Most forensic anthropologists have studied human gross anatomy, usually in a medical school, and all have spent years poring over minute details of the skeleton. The anthropologist may not always be able to identify the species from a fragment of bone, but in most cases, he or she can determine rather quickly if it is human.

Because of their skeletal structure, some species are more difficult than others to distinguish from human remains. In North America, for example, bear bones can be particularly vexing (**Fig. 1**). For people without specialized training or experience it can be very difficult to distinguish between the bones of many mammals and even birds and those that are human.

On occasion, bone fragments are so small that species identification by gross morphology is not possible. In some of these cases, a microscope may be helpful. Because of the way in which the mineral component of bone (mainly calcium and phosphate) is deposited and later remodeled, the microscopic structure or (histomorphology) of bone is unique. And, since the bone of different species is deposited in different patterns it is usually possible to identify human bone fragments from microscopic sections. **Figure 2** is a comparison of a cross-section of human

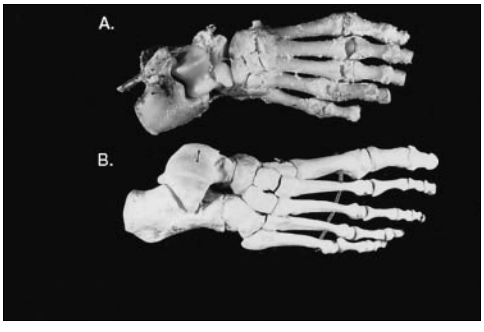


Figure 1 (**A**) Bear foot (**B**) human foot. The human foot is a right foot; the bear foot is a left foot. Both are viewed from above.

A

B

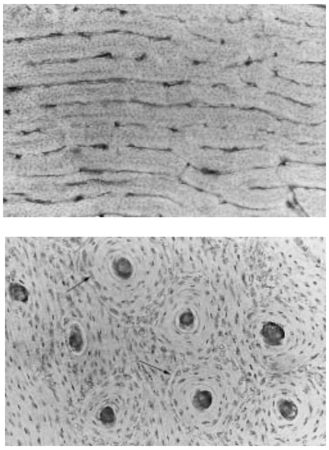


Figure 2 Microscopic sections of (**A**) cow and (**B**) human cortical bone. The arrows indicate Haversian systems in the human bone.

bone and one from a cow. Note that the human bone displays a number of circular structures, known as Haversian systems, distributed among a backdrop of parallel layers. Haversian systems result from the remodeling of parallel layers that are generally concentric to the cross-section of bone. The rectangular (plexiform) pattern displayed by the cow section is generally characteristic of herbivores. There are a few species where confusion with humans may arise, large dogs or wolves, for example, but to an experienced bone histomorphologist, human bone is unique.

Are the remains recent?

Estimating the time since death is one of the most difficult tasks of the forensic anthropologist. When a fresh human body is deposited in the environment, a series of regular and relatively rapid changes begin to occur. If the remains are discovered within a few months, it is often possible to estimate the post-mortem interval to within a few weeks. In such cases input from a forensic entomologist, who may analyze insects associated with the corpse, may be very helpful. After the soft tissue has decomposed and only the skeleton remains, however, the decompositional process slows down. The bones lose their moisture content and the fatty marrow begins to leach out of the medulary cavity (hollow space inside the bone). The bones become lighter, lose the smell of fat and decomposing organic matter and usually begin to take up stain from the surroundings or soil. After a few years it may be very difficult to distinguish between bone that is generally too old to be of forensic interest (50 years or more) and the remains of a more recently deceased individual that might be identified.

The context of the remains may provide key information for the determination of time since death. Evidence of past burial practices, including ancient artifacts, like stone or early metal tools, may signal a particular cultural horizon and therefore indicate time of death. Even coffin handles or coffin nails may help date remains. Because the direct association between such associated items and human remains must be firmly established, a thorough understanding of the conditions of discovery is essential. Forensic anthropologists often work closely with archaeologists, or may have experience in archaeological techniques themselves. Archaeological methods of recovery have proved invaluable for the collection of decomposed human remains and for understanding their context.

Who is it?

If it can be established that the remains are human and are recent enough in origin to be of medicolegal interest, then identification is likely to be attempted. Identification normally begins with the construction of a biological profile, the primary elements of which are: age, sex, race and height. In some cases evidence for medical history (such as a healed bone fracture), may provide clues to identity. With a biological profile, authorities are able to begin scanning their missing person reports and hopefully provide one or a few candidates for positive identification. If such a profile does not yield a positive identification, a facial approximation (from the skull) may be carried out.

Estimation of age Estimating age at death can be extremely difficult, especially as people mature. Although it may be possible to estimate the age of a child or infant to within a few months, the age ranges for adults may be 20 years or more. Because different methods are used to estimate the age of infants and children, young adults and older adults, the methods for each are discussed separately.

Infants and children (fetal to about 15 years) The best methods for estimating fetal age and that of young children involve dental development. Deciduous (baby) teeth begin to calcify early in the fetal period and continue developing until the third year, at which time the development of the permanent dentition is well underway. By comparing to established norms, the stage of formation of each available tooth and the eruption of teeth to the chewing surface, biological age can be estimated to within several months or a couple of years (**Fig. 3**). Because dental enamel is the hardest substance in the human body and, therefore, most likely to be preserved, dental development is particularly useful for forensic cases.

Less accurate than dental development, but also important for estimating skeletal age is bone growth

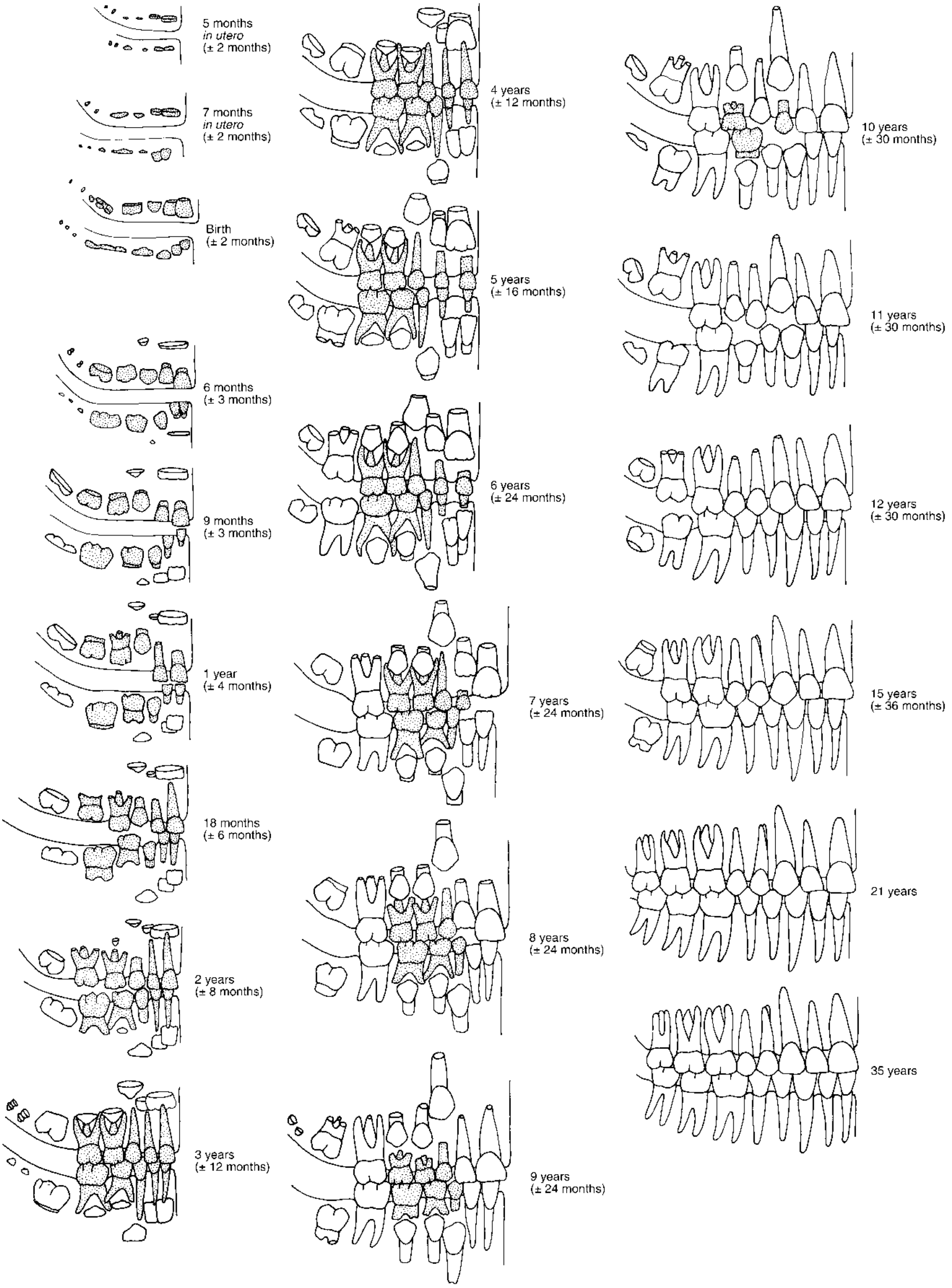


Figure 3 The sequence of formation and eruption of teeth among American Indians (Adapted from Ubelaker 1989, p.64).

and development. In cases involving fetal remains and those of young children, it is often possible to estimate age by comparing the length of long bones (like the femur or tibia) to published charts. Because rates of growth vary in different groups, it is important to use data collected from a population that is most representative of the remains in question.

Finally, the appearance and union of growth centers can be useful tools for subadult age estimation. Most bones do not grow as single units, but as primary and secondary centers of growth or ossification. The primary center is the first to develop, the shaft of the femur, for example, at around the eighth-week of fetal life, and the secondary centers at either end (epiphyses) appear much later in fetal development or after birth. Bones with multiple centers of ossification cease to grow when the primary and secondary centers of ossification fuse together. Because the various epiphyses of different bones appear and fuse at different times, the study of growth centers is a valuable tool for estimating the age of children and young adults. Most useful for estimating the age of older children and young adults are the femur, the shoulder, the clavicle and portions of the hip bone.

Young adults (15 to about 35 years) Epiphyseal union continues to be an important age indicator in young adults. The last epiphysis to completely fuse is the medial end of the clavicle, usually at between 25 and 30 years of age.

Beginning in the later teenage years, a number of systematic changes have been recognized to occur on several joint surfaces. Probably the best known is the metamorphosis of the pubic symphysis (the joint between the two pubic bones of the pelvis). Initially the surface of the joint is characterized by a series of deep horizontal furrows and the joint flares anteriorly. During the twenties the furrows fill in and the build-up of bone causes the opposing bone surfaces to flatten. Usually during the thirties, a recognizable rim forms around the margin of the joint. The transformation of the pubic joint surfaces begins at about 18–20 years of age and continues throughout life (**Fig. 4**). Age can be estimated by comparing the pubic joint surface to available descriptions, illustrations and casts. For a variety of reasons, including the involvement of the pubic joint in childbirth, pubic symphysis aging is most reliable in males and, for both sexes, the reliability of pubic symphysis declines dramatically after about 35 years of age.

Regular changes have also been recognized for the sternal (chest) end of ribs. In the 1980s, Susan Loth and Yasar İşcan began to collect and analyze the sternal ends of fourth ribs from a large sample of individuals of known age and sex. They described for males and females, how the sternal extremity of the rib changes from a flat, billowed surface to first a 'V' shaped then a 'U' shaped pit. As the pit deepens, its surface becomes pitted and the margins become increasingly irregular. As is the case for the pubic symphysis, photographs accompany the description of the changes in the sternal extremity of the fourth rib. Also, like the pubic symphysis, the rib ends become increasingly unreliable after the third or fourth decade. An advantage of both these methods is that they involve bones that are close to the surface and can easily be removed during autopsy.

Older adults (over 35 years of age) Beginning during the fifth or sixth decades of life, a number of 'degenerative' changes may be identifiable on the skeleton. Joint surfaces often begin to show progressive changes associated with arthritis. Tiny spicules of bone may appear on the margins of the vertebral bodies (osteophytosis) (**Fig. 5**) and some lipping may appear around the edges of moveable joints like the knee and elbow or hands and feet. The pubic symphysis and rib ends also show signs of advancing age.

As individuals mature, the dentition and supporting bone also display characteristic changes. The chewing surfaces of teeth slowly wear, the effects of dental disease accumulate, and the bone of the maxilla and mandible tends to resorb. Older people are also more likely to be missing certain teeth.

Aging by histomorphology It was mentioned above that the microscopic structure of bone is useful for distinguishing between human and nonhuman bone. Because of changes in microstructure that occur during an individual's life (for example, Haversian system accumulate), it is also possible to estimate the age at death of an individual by studying sections of bone under a microscope.

Determining sex Because most sexual dimorphism (differences in shape between males and females) develop during puberty, most forensic anthropologists agree that the determination of sex from the skeleton is only practical for late teenagers or adults. Generally, male skeletons are larger and more robust than female skeletons. By studying features of the skull and especially the hip bones, experienced osteologists should be able to identify sex with a better than 95% degree of accuracy.

Innominates (hip bones) Three features on the pubic region of the innominates are particularly useful indicators of sex; the ventral arc, the subpubic concavity and the shape of the ischiopubic ramus

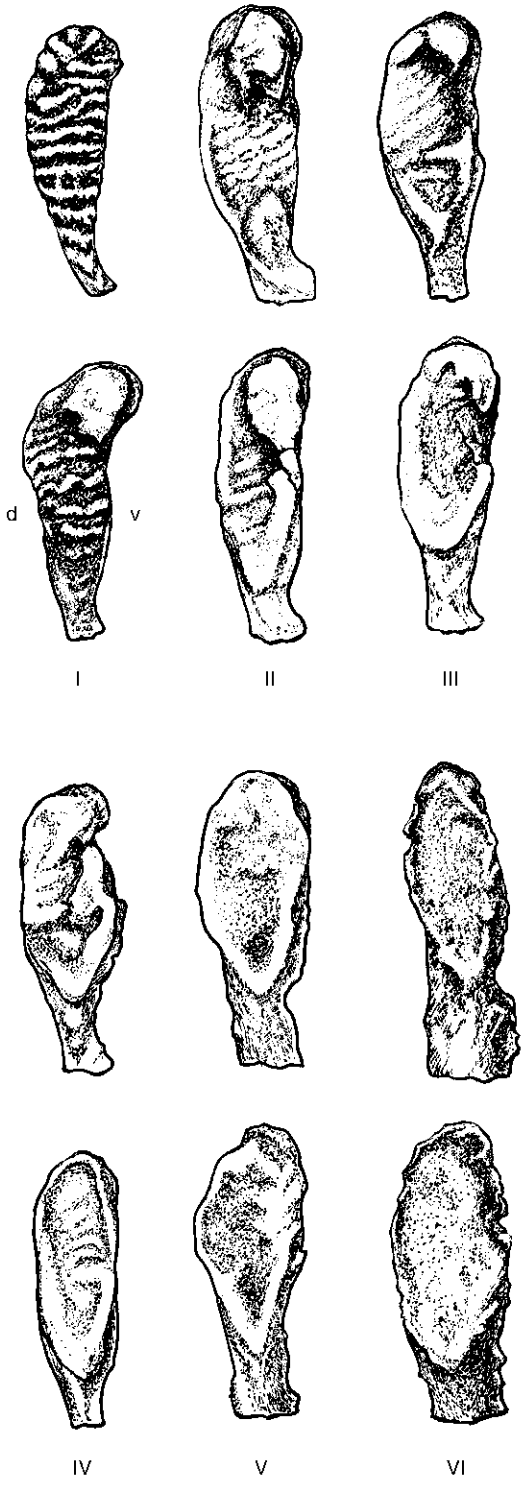


Figure 4 Suchey–Brooks pubic symphyseal phases (I–VI). V, ventral side. Redrawn from Burch and Suchey, 1986

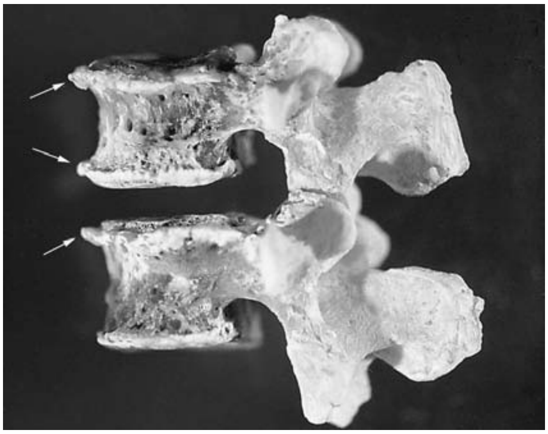


Figure 5 Two lumbar vertebrae of a 60-year-old male. Arrows indicate osteophytosis.

(**Fig. 6**). The ventral arc is a curved ridge of bone on the anterior surface of the pubic bone just lateral to the pubic joint. It is a common feature in females and almost never seen in males. Viewed from the front, the medial edge of the bone that extends laterally and inferiorly from the joint between the pubic bones is often concave in females and straight or convex in males. And finally, the medial (inner) edge of the bone connecting the pubic bone and the ischium is normally wide and flat in males and a narrow ridge in females. If a pubic region displays a ventral arc, subpubic concavity and narrow ischiopubic ramus, the specimen is very likely to be female. Another sign that a hip bone is from a female is a wide subpubic angle (Fig. 6).

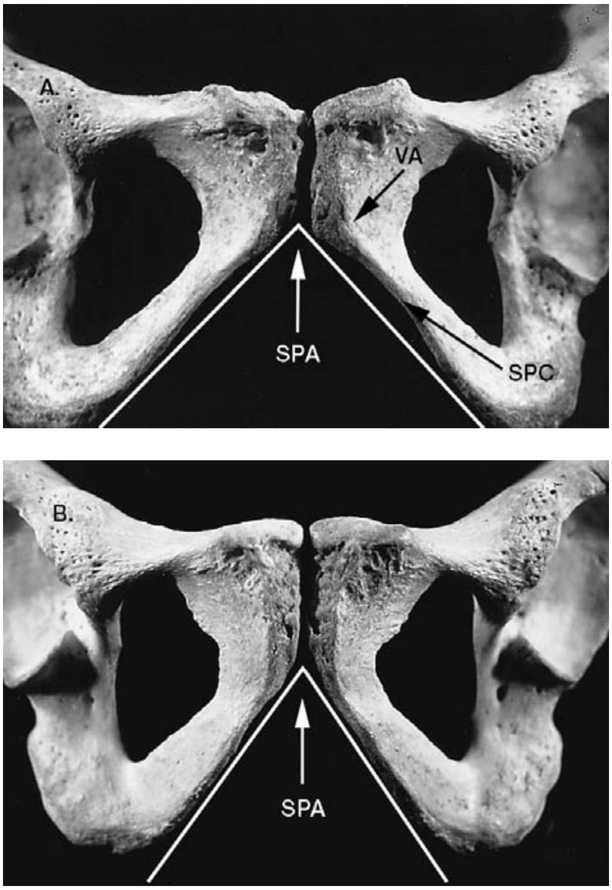


Figure 6 Female (**A**) and male (**B**) pubic bones. VA, ventral arc; SPC, subpubic concavity; SPA, subpubic angle.

The skull Sexually dimorphic features of the skull include the brow ridges, the mastoid processes, and the external occipital protuberance, all of which are usually more pronounced or larger in males (**Fig. 7**). The chin portion of the mandible tends to be squarer in males and more rounded (or 'V' shaped) in females.

Ancestry and race The estimation of one's ancestry is complicated by the fact that most anthropologists agree that dividing the World's human populations into three or four broad races is unscientific and not warranted on biological grounds alone. Nonetheless, the popular notion that there are three or four races of humankind permeates much of our thinking and is the framework usually used by the law enforcement community to identify missing persons and suspects. Therefore, it is usually useful to include in a biological profile, a prediction about the racial identity of a set of unidentified human remains. This is usually accomplished by identifying the broad geographic region of one's ancestry and translating that into terms that are used in the community. In the US, remains that suggest European ancestry are likely to be labeled 'White'.

To estimate ancestry, investigators usually evaluate features of the skull like the shape of the eye orbits, the nasal opening, and the zygomatic or cheek bones. Some anthropologists find the root of the nasal

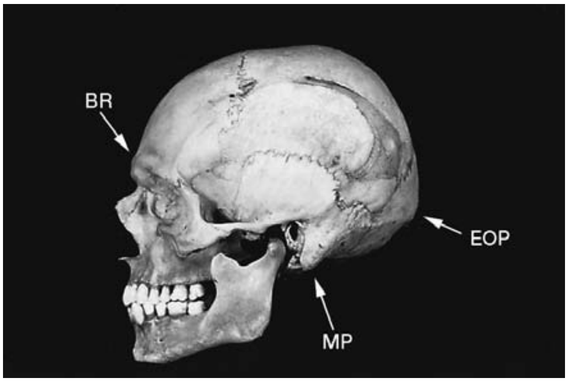


Figure 7 Male cranium. BR, brow ridge; MP, mastoid process; EOP, external occipital protuberance.

bones in the area between the eye sockets particularly useful. Skull shape may also be evaluated by taking a series of measurements. A computer program may then be used to evaluate ancestry by comparing cranial dimensions of an unidentified specimen to those from a large sample of crania of known ancestry.

Stature Anthropologists usually use one of two methods to estimate living height from the skeleton: the anatomical method or limb bone proportionality. The anatomical method involves measuring all of the skeletal elements that contribute to height and then adding a constant. To use the method, the investigator measures the distance from the top of the skull to the foramen magnum, where the skull joins the uppermost vertebra; the height of all of the cervical, thoracic and lumbar vertebral bodies; the femurs; the tibiae; and the talus and calcaneus of the foot. A constant of 10–12 cm is then added to account for soft tissue and other distances that are not taken into account by the measurements.

The limb bone proportionality method relies on well-studied relationships between the length of limb bones and living height. The formulas used most often today are equations that were originally developed in the 1950s on documented anatomical collections and victims of World War II and the Korean War. Because of the variation in long bone-to-height ratios among the world's populations, different formulas are provided for different groups. Probably because of ease of use and the fact that the skeleton need not be so complete, forensic anthropologists rely more heavily on the limb bone proportionality method for estimating living stature from the skeleton.

Secondary features (SES, medical history, markers of habitual activity) In addition to the four customary categories of most biological profiles (age, sex, ancestry/race and height) forensic anthropologists may also contribute information about socioeconomic status, medical history and evidence for habitual activity. Although such generalizations must be approached with caution, it is sometimes possible to estimate that a person is from a lower or higher socioeconomic level because of the location of the remains and associated items like clothing or jewelry. Evidence for certain dental procedures or lack thereof may also indicate access to particular types of treatment. A skeleton that presents advanced untreated dental caries (cavities) or evidence for particularly severe and chronic gum disease may signal that a person is from a low socioeconomic status.

Certain aspects of medical history may also be useful for identification. For example, the presence of an artificial limb or joint may help match an unidentified skeleton with a missing person. In one recent case in our lab, an unidentified skeleton was diagnosed with a rare metabolic disease, acromegaly, which dramatically affects the growth of bone. If a missing person were described as suffering from acromegaly, then the recognition of the disease on the skeleton might have been very useful for identification.

Throughout an individual's life, the bones of the skeleton continue to remodel and reshape themselves according to stresses that are placed upon them. This is known as Wolff's law. By studying the robustness of selected bones and the nature of certain muscle attachments it is sometimes possible to draw conclusions about a person's activities during life. A person who spent most of their life as a stone mason, lugging around cement blocks, is likely to have a very different looking skeleton than a typical anthropology professor.

Facial approximation A useful addition to the biological profile (age, sex, height, ancestry, etc.) is often a facial approximation. Traditionally, facial approximations entail placing tissue depth markers at a number of specified points on and around the facial skeleton, then connecting these markers with clay. Nose, lips and ears are formed using certain skull and tooth dimensions as guidelines. Prosthetic eyes are carefully positioned in the eye sockets before the surrounding eyelids are shaped. The final product (**Fig. 8**) is an image that can be distributed to regional law enforcement agencies and displayed on television and in the newspapers. It is hoped that someone will recognize the face and contact the authorities.

Figure 8 Facial approximation from the skull of a female believed to have been 30–50 years of age at the time of death.

In recent years, facial approximations are increasingly being carried out with the aid of computers. An image of the skull is displayed on the screen and the soft tissues are added as appropriate for the biological profile. Some experts prefer to sketch soft tissue features over an image of a skull, then scan images of both the skull and drawing onto a computer screen for final adjustment. Computer assisted approximations are not only quicker to carry out than clay reconstruction, they are also easier to modify after the process is complete. There are no reliable studies that compare the accuracy or effectiveness of the various methods of facial approximation.

Positive identification Once a biological profile and possibly a facial approximation have been generated for a set of remains, they can be compared to lists and descriptions of missing persons. If there is a match in the general features, then an attempt is normally made at positive identification. Most commonly, positive identifications are made by comparing X-rays (usually dental) taken during a missing person's life with X-rays taken from the same area of the unidentified skeleton. If, in the judgment of an expert with training and experience in the interpretation of X-rays, there is a match in a sufficient number of features, then the remains may be considered positively identified. **Figure 9** is a comparison of the frontal sinuses of a missing male, and of a skeleton that was discovered in a nearby wooded locale. X-rays of parts of the skeleton other than the skull, such as limb joints and the vertebral column, are also useful for positive identification.

Sometimes, when there are no appropriate antemortem (before death) X-rays available for a missing person, it may be necessary to compare the shape of the skull and its features to a photograph taken during life. This process, known as skull–photo superimposition, may, under certain circumstances, also yield a positive identification. Skull–photo superimposition, is performed more often in parts of the world, e.g. China and Japan, where antemortem X-rays are less common.

Manner and Cause of Death

In addition to identifying human remains, forensic anthropologists are often called upon to help interpret evidence for manner or cause of death.

Marks on the bones themselves can provide very important clues about how a death occurred. Bullets penetrating the skull, for example, typically create a characteristic defect (**Fig. 10**). From such damage, an expert may be able to make judgments about the maximum caliber of a gun and the angle and velocity of a projectile. Similarly, details of weapons may be indicated by blunt force damage (like that from a hammer) or various knife strikes. It has shown that it may even be possible to identify a specific knife using electron microscope images of skeletal damage.

With all evidence of skeletal trauma, it is imperative to recognize and distinguish among antemortem (before death), perimortem (around the time of death) and postmortem (after death) trauma. A skull fracture may indeed reflect a blow to the head that caused death. But before such a conclusion is

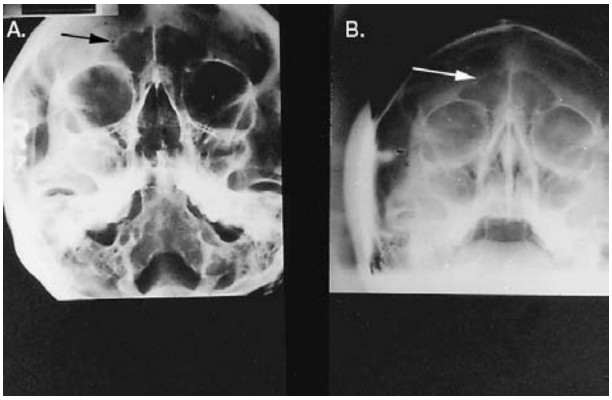


Figure 9 Comparison of postmortem (**A**) and antemortem (**B**) images of the frontal sinus of a male skull (arrows).

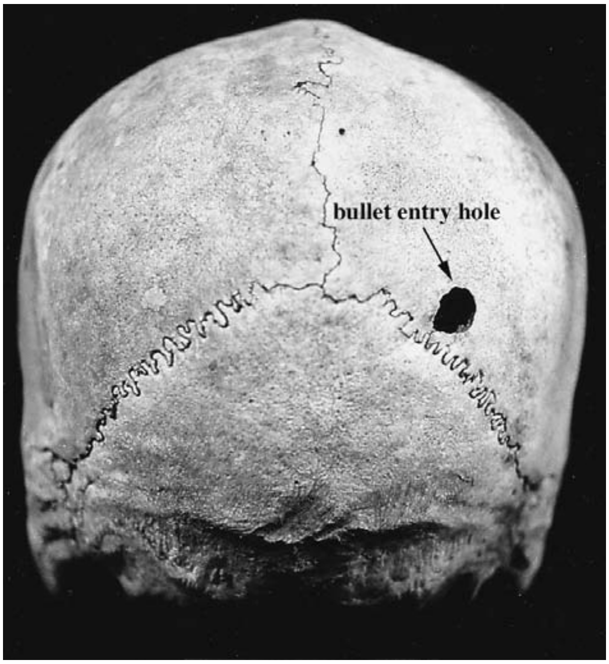


Figure 10 Bullet damage to the rear of a skull. The opening in the inside is much larger and reflects a cone-shaped defect.

drawn, it must be established that the damage is not evidence of an old injury, from which the victim survived, or that the damage occurred to the skull as it lay on the forest floor long after death. The hallmark of antemortem injuries, evidence for healing near the margins of the fractured bone, may appear as soon as two or three weeks after an injury (**Fig. 11**).

Because dried bone has different fracture characteristics than bone from living organisms, it is sometimes possible to distinguish between perimortem and postmortem damage by studying details of skeletal lesions. Differential staining may also indicate postmortem damage. When bone lies in a soil environment for a period of years, the exposed surfaces typically become stained from ground water. When a fracture occurs at or near the time of discovery, a new unstained surface may be created that contrasts distinctly with the stained bone.

The context in which remains have been recovered can also provide important clues about how a person dies. The techniques of forensic archaeology may demonstrate that the orientation of a body (for example, with the hands bound behind the back) or the position of a potential murder weapon are clues that a homicide took place and that death was unlikely to have been accidental.

The Forensic Anthropology Laboratory

Although all forensic anthropologists are capable of generating a biological profile from a person's skeletal remains and assisting with positive identification, some have specialized in specific areas like trauma analysis, facial approximation or microscopy. Similarly, forensic anthropology laboratories are variously equipped to provide different kinds of services. Presumably a forensic anthropology lab would minimally include adequate work surfaces for the examination of a set of skeletonized human remains, appropriate water sources and disposal facilities, a comparative skeletal collection and a full range of calipers. An ideally suited forensic anthropology lab might also include facilities for handling remains in all stages of decomposition; instrumentation for thin section preparation and microscopic analysis, including electron microscopy; image superimposition equipment; and instrumentation for DNA analysis. Although such labs and combinations of expertise are rare, cooperation among forensic anthropologists and other scientists usually allows for sufficiently complete anthropological evaluations of cases involving decomposed remains.

Figure 11 The right side of a fractured human jaw. Notice the evidence for healing (arrows).

See also: **Accident Investigation**: Motor Vehicle; Rail; Determination of Cause: Overview. **Anthropology**: Archaeology; Morphological Age Estimation; Sex Determination; Determination of Racial Affinity; Excavation and Retrieval of Forensic Remains; Bone Pathology and Antemortem Trauma in Forensic Cases; Skeletal Trauma; Animal Effects on Human Remains; Assessment of Occupational Stress; Stature Estimation from the Skeleton. **Causes of Death**: Overview; Scene of Death; Postmortem Changes; Sudden Natural Death; Blunt Injury; Sharp Injury; Traffic Deaths; Systemic Response to Trauma. **Crime-scene Investigation and Examination**: Recording. **Deoxyribonucleic Acid**: Significance. **Entomology**. **Identification/Individualization**: Overview and Meaning of ID. **Odontology**. **Pathology**: Overview. **Accident Investigation**: Airbag Related Injuries and Deaths; Driver Versus Passenger in Motor Vehicle Collisions.

Further Reading

Bass WM (1987) *Human Osteology: A Laboratory and Field Manual*, 3rd edn. Columbia: Missouri Archaeological Society.

Burch M, and Suchey J (1986) Age determination of the male OS pubis. *Physical Anthropology* **69**:427–435.

Galloway A, Birkby WH, Kahana T and Fulginiti L (1990) Physical anthropology and the law: legal responsibilities of forensic anthropologists. *Yearbook of Physical Anthropology* **33**:39–57.

Haglund WD and Sorg MH (1997) *Forensic Taphonomy: The Postmortem Fate of Human Remains*. Boca Raton: CRC Press.

Hunter JR and Roberts C (1996) *Studies in Crime: An Introduction to Forensic Archaeology*. London: Batsford.

Houck MM (1998) Skeletal trauma and the individualization of knife marks in bones. In: Reichs K (ed.) *Forensic Osteology: Advances in the Identification of Human Remains*, 2nd edn. Springfield: Charles C.

İşcan MY and Kennedy KAR (1989) *Reconstruction of Life from the Skeleton*. New York: Alan R. Liss.

Rathbun TA and Buikstra JE (eds) (1984) *Human Identification*. Springfield: Charles C. Thomas.

Reichs KJ (1998) *Forensic Osteology: Advanced in the Identification of Human Remains*, 2nd edn. Springfield: Charles C. Thomas.

Stewart TD (1979) *Essentials of Forensic Anthropology: Especially as Developed in the United States*. Springfield: Charles C. Thomas.

Ubelaker DH (1989) *Human Skeletal Remains: Excavation, Analysis and Interpretation*, 2nd edn. Washington, DC: Taraxacum.

Skeletal Trauma

M Y İşcan, Adli Tıp Enstitusu, Istanbul Univesitesi, Istanbul, Turkey

G Quatrehomme, Faculté de Médecine de Nice, Laboratoire de Médecine Légale, Nice, France

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0760

Introduction

One of the purposes of the medicolegal system is to protect the innocent and prosecute the guilty. Evidence is needed to try the accused. An analysis of trauma that may have caused death to the victim is an important step in this direction. It may also give information about antemortem injuries such as those resulting from accidents, child abuse, torture or other human rights violation. The term trauma may be defined in a number of ways. However, it is the response of the body against any sudden external insult. It may alter an individual's life from normal to a functional and esthetic abnormality and may lead to death or irreversible injury to the victim. Determination of the cause of trauma is a difficult task and therefore may be very complex to explain. The problem may be worse if the remains are skeletonized or in a state of advanced decomposition.

Anthropological literature classifies trauma by its origin into dental trauma, fractures, dislocation, scalping, surgery, perimortem cuts and breaks, and weapon wounds. Some of these categories are primarily antemortem in origin (e.g. dental trauma, fractures, dislocations) and may not be responsible for acute death. Some (e.g. surgery) are handled in hospitals. Of the remaining, perimortem cuts and breaks and weapon wounds are of immediate concern for forensic scientists, particularly if the victim died because of them. Once the trauma is classified as such it must be assessed in terms of its relation to death. So, a trauma can be considered antemortem, perimortem or postmortem. Trauma-like appearance of the skeleton may also be caused by other agents such as animal or other environment events.

No matter what the origin of trauma it must be assessed with suspicion within the perspective of legal sciences. The ultimate aim of such an assessment is to determine not only the cause of the damage but also the manner of death. The purpose of this article is to discuss the causes of trauma that are observable from the skeleton and their implication in forensic sciences.

Cause and Manner of Death

In the case of a crime involving injury or death, the cause must be established. If the victim dies from it, the manner must be decided. Terms cause and manner have specific meanings in forensic medicine. The cause may be defined as an injury or disease which may lead to death. Gunshot wound to the head is a cause that will often kill the victim.

Manner is the classificatory system of death and can be one of four main and classical assessments: natural, accidental, suicide and homicide. The decision as to which one of these is to be considered as the manner is made by the authorized individual, e.g. medical examiner or coroner. The coroner (or medical examiner) in turn needs the advice and evidentiary support from the pathologist, anthropologist, criminalist and others, who are involved in the investigation, and have expertise of the crime and scene where the deceased is found, and have expertise to explain what may have killed the victim. Using the gunshot wound as the cause of death as an example, the manner can be one of several. The victim may have killed himself/herself with a single shot into the mouth and thus it may be considered suicide. The person may have been shot from the back of the head several times by an assailant. This suggests homicide. The gun may have been accidentally fired while cleaning. Natural deaths may occur for a number of reasons including malfunctioning of vital organs such as heart attack and other acute diseases, or terminal death resulting from chronic diseases.

In order to make decisions about the manner, a number of issues must be resolved. There are two kinds of remains that may shed light into the cause of death. The first, includes those that have an obvious

İşcan MY and Kennedy KAR (1989) *Reconstruction of Life from the Skeleton*. New York: Alan R. Liss.
Rathbun TA and Buikstra JE (eds) (1984) *Human Identification*. Springfield: Charles C. Thomas.
Reichs KJ (1998) *Forensic Osteology: Advanced in the Identification of Human Remains*, 2nd edn. Springfield: Charles C. Thomas.
Stewart TD (1979) *Essentials of Forensic Anthropology: Especially as Developed in the United States*. Springfield: Charles C. Thomas.
Ubelaker DH (1989) *Human Skeletal Remains: Excavation, Analysis and Interpretation*, 2nd edn. Washington, DC: Taraxacum.

Skeletal Trauma

M Y İşcan, Adli Tıp Enstitusu, Istanbul Univesitesi, Istanbul, Turkey
G Quatrehomme, Faculté de Médecine de Nice, Laboratoire de Médecine Légale, Nice, France

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0760

Introduction

One of the purposes of the medicolegal system is to protect the innocent and prosecute the guilty. Evidence is needed to try the accused. An analysis of trauma that may have caused death to the victim is an important step in this direction. It may also give information about antemortem injuries such as those resulting from accidents, child abuse, torture or other human rights violation. The term trauma may be defined in a number of ways. However, it is the response of the body against any sudden external insult. It may alter an individual's life from normal to a functional and esthetic abnormality and may lead to death or irreversible injury to the victim. Determination of the cause of trauma is a difficult task and therefore may be very complex to explain. The problem may be worse if the remains are skeletonized or in a state of advanced decomposition.

Anthropological literature classifies trauma by its origin into dental trauma, fractures, dislocation, scalping, surgery, perimortem cuts and breaks, and weapon wounds. Some of these categories are primarily antemortem in origin (e.g. dental trauma, fractures, dislocations) and may not be responsible for acute death. Some (e.g. surgery) are handled in hospitals. Of the remaining, perimortem cuts and breaks and weapon wounds are of immediate concern for forensic scientists, particularly if the victim died because of them. Once the trauma is classified as such it must be assessed in terms of its relation to death. So, a trauma can be considered antemortem, perimortem or postmortem. Trauma-like appearance of the skeleton may also be caused by other agents such as animal or other environment events.

No matter what the origin of trauma it must be assessed with suspicion within the perspective of legal sciences. The ultimate aim of such an assessment is to determine not only the cause of the damage but also the manner of death. The purpose of this article is to discuss the causes of trauma that are observable from the skeleton and their implication in forensic sciences.

Cause and Manner of Death

In the case of a crime involving injury or death, the cause must be established. If the victim dies from it, the manner must be decided. Terms cause and manner have specific meanings in forensic medicine. The cause may be defined as an injury or disease which may lead to death. Gunshot wound to the head is a cause that will often kill the victim.

Manner is the classificatory system of death and can be one of four main and classical assessments: natural, accidental, suicide and homicide. The decision as to which one of these is to be considered as the manner is made by the authorized individual, e.g. medical examiner or coroner. The coroner (or medical examiner) in turn needs the advice and evidentiary support from the pathologist, anthropologist, criminalist and others, who are involved in the investigation, and have expertise of the crime and scene where the deceased is found, and have expertise to explain what may have killed the victim. Using the gunshot wound as the cause of death as an example, the manner can be one of several. The victim may have killed himself/herself with a single shot into the mouth and thus it may be considered suicide. The person may have been shot from the back of the head several times by an assailant. This suggests homicide. The gun may have been accidentally fired while cleaning. Natural deaths may occur for a number of reasons including malfunctioning of vital organs such as heart attack and other acute diseases, or terminal death resulting from chronic diseases.

In order to make decisions about the manner, a number of issues must be resolved. There are two kinds of remains that may shed light into the cause of death. The first, includes those that have an obvious

cause of death, e.g, a bullet hole in the head. This type of case is not considered circumstantial. It is well known that gunshot wound to the head may kill (and very often it does kill) a person, and the wound has particular entrance and exit characteristics. In the second type there is damage or injury to the skeleton, e.g. marks on the ribs or a broken zygomatic bone. In this case the cause of death may be considered circumstantial, because it is not definite that these marks caused the death. The damages may have been made by a knife, an animal scavenging or something else, and the death may have occurred from another cause.

To determine if the cause is obvious or circumstantial, the crime scene must be very carefully investigated and must not be contaminated by factors such as stepping on evidence or the relationship of scattered artifacts and remains with each other. Forensic techniques, and particularly forensic anthropological (archeological) techniques are essential to analyze a crime scene. It is also important to positively identify the victim in order to solve a circumstantial death case. In situations where the scene is the remains of a highly decomposed person, it is of benefit to the coroner to invite a forensic anthropologist who is competent in archeology and osteology. A forensic anthropologist is a consultant to the legal system and commonly involved in such crime scene assessment and identification of the deceased. In order to determine whether a particular trauma is death related or antemortem or what the manner of death is, his/her assistance is essential.

In general there is a tendency by crime scene police to underestimate the importance of a crime scene where skeletonized remains are found. And yet it is ironic that the same experts are very meticulous when they deal with the scene where a recent crime has occurred.

Wound Assessment

Antemortem trauma may be obvious when there is healing, inflammatory process, or periosteitic/osteomyelitic activity at the location of the lesion. It is well known that the inflammatory process starts very soon after an injury, and may be visible within a few hours, at least at the microscopical level. Bone inflammation, such as periosteitis and osteomyelitis, is a gradual process and therefore chronic, and the presence of bone reaction and remodeling of the injury indicates that a time period of several weeks or months may have passed since death.

Wound assessment also gives clues in child abuse cases. The victims of this often show healed lines of fracture in their bones, particularly in the ribs. Although various stages of healing have been described, there are many factors that may complicate the problem. These include age, sex, the bone itself (rib, skull, or long bone), the nature of the instrument causing the wound, and even the general health of the victim (e.g. rickets).

There are a lot of pitfalls to be avoided. A healed fracture is obviously an antemortem condition and the person lived some time for the wound to heal. A fresh wound that may have occurred a few weeks before death is thus not an immediate cause of death. Perimortem trauma refers to a wound that may have occurred at the time of death, or a short time before or after death. In perimortem trauma, it is often impossible to determine if the fracture or defect occurred exactly at the time of death. The forensic scientist has to find some clues, such as the presence of hematoma or absence of discoloration at the edges of the fractured area, and other classic perimortem patterns (e.g. gunshot wounds, patterned blunt or sharp force injuries, butterfly fractures).

Hematoma shows the discoloration around the edges of a broken area. Its assessment may be difficult if there is a long time interval since death. The elapsed time not only affects the color of the hematoma which will gradually fade, but also changes the color due to surrounding material such as the decaying leaves or the soil itself. Nevertheless discoloration in relation to bloodstain of a perimortem fracture is usually very localized, whereas the dark staining by the soil or rotting leaves should be visible not only on that bone but others which are in contact with the plants and ground.

There are several instances in which the problem of wound assessment can be simpler to handle. Recognition of the bleeding of the fractured edges such as that in a hyoid bone may be all that is needed to determine the cause and manner of death. Ecchymosis of the skull is another relevant clue. In this case, blood coming from the broken tiny trabeculae may infiltrate the diploe of the skull vault. This may be readily observed by transillumination of the skull, that is, holding the skull in front of a simple light source, and by microscopy.

Postmortem damage can also be very difficult to analyze. Ribs can easily be broken by a crushing injury or broken by a scavenging dog. An injury caused by a blow to the chest may not be easily differentiated from that caused by a dog or large animal stepping on the remains. Postmortem defects or fractures are very frequent. They are caused mainly by environmental factors, including soil chemical and mechanical erosion, exposure to sun and water, plant activity, and animal activity. Rarely the cause of postmortem alteration is a postmortem trauma due to human activity.

A careful crime scene investigation can give some answers to the problem. It provides important clues about the damaged bones and whether remains were scattered when they were discovered. In this way the footsteps of a cow or horse can be differentiated from a blow on the chest by a baseball bat, metal pipe, or some other blunt object.

Wound Classification

Trauma can be classified into gunshot wounds; sharp force injuries; blunt trauma; and burnt bones. In fact the various trauma encountered routinely are a wide range of combination of these four elementary lesions. Of these the first three are made by an instrument or a weapon, and sometimes by the head, hands or feet of the assailant.

Burnt bones are relatively simple to recognize (but their forensic analysis raises a great deal of issues) and are described elsewhere.

Gunshot wounds

Most of the research on trauma has been the assessment of gunshot wounds. A gunshot wound has a number of characteristics. The most obvious ones are the entrance and eventual exit holes. Each hole in a skull has a unique appearance of outer and inner tables (external and internal openings) of an entrance hole.

The exit holes also have internal and external openings. The exit hole is morphologically different from the entrance. The entrance is typical when it is round or ovoid shaped, sharp-edged of punched-out appearance, and interiorly beveled. The exit is highly varied depending upon which bone the bullet exits and whether the bullet is jacketed or not, but is usually larger than the entry wound, and more irregular. Jacketed bullets should make less damage to the skull. If the exit is through one of the thin bones as those of the face, the whole area may collapse and there may be no sign of which of the bones was hit first and could become very difficult to interpret.

It should be noted that exit and entry wounds are better known and thus assessed for the skull than for the postcranial skeleton. This is probably because of the skull's anatomical shape (a globe with two tables separated by the diploe). Even large bones like the pelvis and scapula do not show the same gunshot wound opening characteristics as those of the cranium.

Tangential shots are those that make a grazing defect on the surface of the bone and then deflect away. This occurs only because the angle of trajectory is very small. These wounds are relatively easy to recognize on the surface of the skull or postcranial skeleton. However, the skull wounds are more complex and varied. Tangential wounds may vary from a simple grazing creating an oval-shaped nick (e.g. on the skull) or a semilunar defect (e.g. on the edges of a rib). Sometimes a part of an irregular bone is broken off. This category can be exemplified with the breakage of the transverse process of a vertebra or the zygomatic process of the temporal bone. The bullet may hit the surface of a bone with a very narrow angle and thus continue grazing until it forms a groove. This may occur more commonly in bones with a flatter surface like the bones of the skull vault or the mandibular ramus. Although these three types of tangential wounds may be seen in the entire skeleton, the skull shows further variations which usually are considered as guttered and keyhole defects. The skull shows an outer and inner table separated by diploe. Because of this the wound formation varies. In essence a tangential wound may be seen only on the outer table. Sometimes both tables are affected. However, the inner table lesion may be due to the pressure arising from the impact on the outer table. In a third case the inner table exhibits a defect, but the bullet exits the skull without a massive fracture.

Keyhole defects are typical but relatively rare. The bullet hits the skull at a shallow angle and splits into two parts immediately. One portion penetrates the skull and explains the occurrence of a typical entry wound with internal beveling. The second portion is deflected and explains the occurrence of an exit wound with external beveling, connected to the entry wound.

The angle of the shooting can also be deduced from the skeleton, the skull being a better indicator than the postcranial skeleton unless several bones (e.g. sternum, ribs and a vertebra) are fractured along the path of the bullet. In the latter case one can determine the direction with a simple knitting pin. In the cranium, the shape of the entrance hole and beveling are of great importance. A circular entry and symmetrical beveling suggests, for example, a right-angle shot. Oval or oblique entry suggests a narrow angle between the direction of the bullet and the surface of the damaged area of the skull. In other cases, though rare, grazing, gutter wounds or keyhole indicate a tangential shot.

The direction of the shot is deduced from the entrance and exit holes as the path is usually straight between them. This direction is described by giving the angles in reference to a sagittal and horizontal (Frankfort horizontal) plan, and the right/left direction, on an individual standing in the standard anatomical position. Radiographs are important, especially

when there is no exit. Autopsy and radiographs can detect the projectile or its fragments within the brain. Metallic ‘snow’ is a classic appearance in the radiographs. Sometimes the pathway is quite amazing. A bullet may follow the curvature of the skull for a long distance, or strike a rib at an angle and encircle the chest, and possibly exiting the opposite side without having penetrating the pleural cavities. Sometimes the bullet penetrates the heart and travels down the femoral artery. These erratic pathways may complicate the interpretation of the death scene, especially from skeletonized remains.

The range of fire is another important issue. Finally the production of fractures is dependent on the range at the time of discharge and the kinetic energy possessed by the bullet, as well as the shape of the missile when it perforates the skull. Fractures are very common in contact wounds in the head, due to the gas expansion. The primary fracture is the hole. The secondary fractures are radial fractures, which try to dissipate the kinetic energy. If this is insufficient, tertiary concentric fractures occur. Multiple shots to the head may be assessed by the arrangement of some fractures that are interrupted by others.

Sharp force injuries

A cut or incised wound is seen when a sharp-edged object (weapon, object, any sharp instrument) is drawn over the skin: the injury is longer than it is deep, and the force required is slight. Exposed portions of the body are often involved: head, neck, arms. In sharp force injuries the sharp tool cuts and divides as it penetrates, whereas in blunt trauma there is a crushing or tearing of the tissues. Incised wounds in bones are usually seen on ribs and vertebra. In ribs they are seen in association with fatal chest stab wounds, and similar injuries of the anterior aspect of cervical bodies can be produced by deep slashes on the upper extremities. In hands they are often interpreted as defense lesions. The length of the lesion is really more than the depth; usually the cutting into bone is very superficial because of its hardness. Incised wounds from edged weapons display a straight and even profile when magnified.

Stab wounds are due to bone percussion, more or less at a right angle, with a sharp or partially sharp pointed object; a pointed instrument penetrates more or less deeply, and the wound is deeper than its length. In bones they can be called punctures. The tip of a sword is a good example, but more common examples include the tip of a knife, scissors, screwdriver, but also glass, nail etc. Most of stab wounds are associated with homicides, and the thorax and abdomen are often targeted. The lesion is roughly linear and deep, resulting in a tiny point depression up to a localized large area of multiple fractures. Sometimes the tip of the blade breaks off and remains embedded in the bone.

Chop wounds are inflicted by heavy and sharp instruments (e.g. axes, machetes, meat cleavers). Most chopping wounds are also seen in murder cases. The chest and abdomen are rarely targeted. In contrast exposed portions of the body are often struck (e.g. head, face, neck, shoulders, and extremities). Very often the death is immediate and due to shock and hemorrhage. The injury has about the same size as the cross-section of the penetrating implement. There is a combination of cutting (sharp force injury) and crushing or compression (blunt force injury) mechanism, therefore the edges are sharp (incised appearance) and irregular (blunt trauma). In bones one can expect a patterned groove with two walls and a floor, or a depression with multiple communitive fractures. The cranial surface may be depressed and display a wound which is almost an exact reproduction of the cross-section of the blade. In limbs a bone break is possible which may indicate the direction of the blow. However, the fact that the assailant, the weapon, and the victim are all independently in motion may give a twisting mechanism which could complicate the interpretation of the lesion.

The results of chopping insults are partial or complete amputations of fingers or toes, hands or feet, fractures of limbs at various levels, severe lesions of joints up to a complete disarticulation, and sometimes a glancing blow if there is a tangential strike of the weapon. Other sharp force injuries exhibit complex lesions: electric saws, serrated knives, propellers and so forth. Saws involve rare situations of postmortem dismemberment and mutilation. Multiple repetitive movements produce peculiar tool marks.

Blunt force injuries

Blunt trauma are usually caused by a direct blow (with a blunt weapon of which the most frequent is a natural object as the fist or a blunt tool), a fall on the ground or on an object, and a compression, as when a part of the body is passed over by a vehicular wheel. The lesion is produced by an impact of the body on a smooth or flat surface, or an impact on the body by a smooth (as a bludgeon) or flat instrument. Therefore a blunt impact tears, shears and crushes. The severity of the lesion depends on several factors, mainly the force delivered by unit of surface and time, and the nature of the surface or weapon.

The mechanisms which cause blunt trauma include compression, tension, bending; and twisting. Most

often all four mechanisms may work together to create the trauma. In compression of a long bone, the force is applied onto the axis. The bone tends to fail in oblique angles. This axial compression gives a 45° angle with the line of fracture. Compression is also very frequent in cancellous bone such as the vertebral body. Tension is the opposite mechanism. The bone is stretched in the direction of the axis, and the bone usually exhibits a transverse line of fracture. Pure forces of tension are rarely seen, except when a tendon or ligament exhibits a strong and fast force (e.g. the malleolus of the ankle). In bending there is tension on the convex side, and compression on the concave side of the bone. A butterfly fracture is not rare in these circumstances. In twisting the force is exerted in a spiral direction, and results in very complex shear-stresses, tension and compression stresses, applied to the bone.

In forensic pathology one distinguishes between direct trauma and indirect trauma. Direct trauma (e.g. by a blow with a stick) has specific characteristics: skin lesions in front of the fracture; transverse fractures, usually at the same level on the two bones of arms or legs. In contrast indirect trauma shows no skin lesion: e.g. in oblique or spiral fractures, if the two bones are fractured, the fractures are not at the same level; usually in these cases the mechanism is proved to be indirect, such as a fall or torsion of a limb.

The recognition of blunt lesion is often possible by a careful analysis. Ragged edges and crushing along the margin of a defect indicate a blunt force injury rather than a sharp force. Sometimes the edges of the bone have to be examined under magnification. On the skull, flakes of outer table are due to the levering action of the implement while it is inserted in the wound, and is seen with a chopping instrument, or very heavy blunt strike which may push the bone plates inwards. Delamination of the outer table is a sign of blunt trauma.

In blunt trauma of the head, the plates of bone produced by the radiating fractures are forced inward, and the eventual concentric fractures are beveled internally (in contrast, in gunshot wounds, the plates of bones are levered externally by the intracranial pressure). The results depend on the rate of loading. A slow rate may give a plastic permanent deformation of the plates of bones. The fractures can be redirected, following the weakest pathways. An interesting feature in skull is the occurrence of a bone patterned lesion (indenting fracture), whereas bursting fractures display radiating and concentric fractures. Ring fractures of the base of the skull are peculiar, usually due to the compression of the skull vertically against the spine. Face trauma is very frequent, especially during traffic accidents. Each bone of the face may be affected. Other bones such as the hip and vertebra have their own mechanism and fracture pattern. In general though, a blunt trauma leaves bruises (even in bones), fractures, and dislocations.

Location of the Trauma

Skull trauma is very particular because of its anatomy (the skull is a closed box, and has two tables separated by the diploe). Some parts of the face act as a fender and other parts as a buttress and are thus able to redirect fractures. This may explain why skull lesions are different from those in other bones. Some of these include: dissipation of the energy by radial and concentric fractures; redirection of fractures by buttressing; the possibility of plastic deformation in slow loading cases; trends to lever the plates of bones inwards in blunt trauma, and outwards in gunshot wounds; the possibility of shattering of the skull into multiple pieces, because of the increase of intracranial pressure; the occurrence of internal beveling of the entry wound, and external beveling of the exit wound (in gunshot trauma); difficulty to interpret some lesions occurring on thin parts of the skull (like the thin plates of the orbits); occurrence of contrecoup fractures, occurring at a location away from the point of impact, due to an undulation of the skull vault.

Postcranial trauma differ from those of the cranium. For example a bullet easily passes through the postcranial body especially the abdomen and chest and yet may leave only a few nicks and cuts and sometimes nothing in the bone. Bullets may create tunnel-like damage in larger vertebrae (lower thoracic and lumbar), but very often a piece of the bone is broken off (e.g. vertebra). A bullet usually breaks a rib into two pieces. Beveling is not a rule but is possible in thick bones like the femur and humerus, and even in vertebrae, ribs, and rims of the orbits or of the zygoma processes. These cases are of great interest because they indicate the direction of the trauma as in the skull. Sharp force injuries may also leave clues on various bones. Cutting or stabbing lesions are important evidence. They may be also analyzed in terms of direction (i.e., from front to back or left to right), when several bones are struck (e.g. an anterior surface of a rib and a vertebra). In these cases one has to reconstruct the pathway and assess which organs or arteries were struck. This can lead to the conclusion that a knife injury probably caused death. Some locations indicate defense wounds (fingers of the hands, or cutting marks on the forearms).

Discussion

Assessment of bone trauma is a very important aspect of forensic pathology and anthropology. On skeletonized remains, soft tissue analysis is not possible, but lesions can persist in bones. Therefore an analysis of bone trauma may lead to an understanding of the cause and manner of death, and to some extent assist the investigator to reconstruct the death scene.

From a practical point of view, the lesions encountered are made by a combination of factors such as gunshots, sharp forces, blunt objects, and sometimes heat such as fire can also insult the bone. Any bone lesion must also be assessed as antemortem, perimortem, or postmortem trauma. Sometimes, this can be very difficult because some lesions can occur both during and after death.

Gunshot wounds are usually recognized by specific features belonging to the entry wound, i.e. round or oval in shape, outer clean and punched-out appearance and, internal beveling. These specific features are frequently seen in skulls, at least in the neural part of the skull where bones are thicker. Exit wounds in a skull are not always seen (if the bullet remains within the brain), if present they are usually larger than entry wounds, and nearly always irregular. Wounds exhibit a wide range of shapes, sizes, and other unusual features. These shapes may be due to the nature of the bone itself (its thickness, softness, disease), the bullet (velocity, range, intrinsic features of the bullet, cross-section as it strikes the bone), the interaction between both (e.g. the angle between the bullet and the bone), and other factors.

Sharp force injuries are divided into cutting, stabbing and chopping lesions. Cutting lesions of bone are very superficial and longer than deep due to a sharp blade. Stabbing in bone leads to a well delimited puncture, with a patterned lesion, displaying the cross-section pattern of the blade which has struck the bone. Chopping wounds, due to heavy implements like an axe, act by a double mechanism (sharp force and blunt force injury). Unusual features are frequent, because of blade types, such as serrated knife, double-edged knife, or more commonly of the various sharp or pointed objects, which are involved in bone lesion. Another difficulty of this analysis is represented by the constant moving of the assailant and/or the victim during the assault.

Blunt force injuries are the most frequent lesions encountered in forensic sciences. By definition the instrument does not exhibit a sharp edge, and the mechanism is only an impression of bone. Cortex is compressed below the injury in blunt trauma: in sharp force injury, the compression is weak and cortex is compressed to the side. The compression increases in chopping wounds, because the two mechanisms, sharp and blunt, are involved. Blunt force results in ragged edges of the lesion. Blunt trauma gives the notions of elastic (reversible) and plastic (permanent) deformation, with the tension and compression sides of the trauma. Tubular long bones behave in a manner resembling tubular beams. Therefore, the recognition of the tension and compression sides helps to assess the direction of the trauma, which is from compression side to tension side. Skull fractures are complex, because the inward displacement of the bone (due to the impact) may induce some outward displacement, and may explain the occurrence of fractures at a distance from the point of impact (at which there is sometimes no damage at all). In the skull, areas of buttressing redirect fractures, whereas cranial sutures and vascular grooves are areas of lesser resistance. The patterned feature of some blunt trauma may help to assess the general characteristics of the weapon, or more precisely the cross-section surface of the weapon. The internal beveling of the concentric fractures in skull are explained by inward forcing of the plates of bones (produced by the radiating fractures). This can help to recognize the direction of the force: for example if the head is shored on one side and struck on the other, internal beveling is seen on both sides, because the direction of force is bilateral. Slow loading involves the bone to fracture in plastic deformation, and the bones remain permanently deformed; this situation is seen when the head is run over by a car tire. As for fractures in gunshot wounds, the rule of crossing determines the sequence of multiple trauma to the skeleton, as the second fracture is interrupted by the first one.

See also: **Causes of Death**: Blunt Injury; Sharp Injury; Systemic Response to Trauma. **Clinical Forensic Medicine**: Evaluation of Gunshot Wounds.

Further Reading

Adelson L (1974) *The Pathology of Homicide*. Springfield, IL: Charles C Thomas.

Davis JH (1983) Principles of evidence interpretation. In: Morse D, Duncan J and Stoutamire J (eds) *Handbook of Forensic Archaeology and Anthropology*, pp. 186–193. Tallahassee: Bill's Book Store.

DiMaio DJ and DiMaio VJM (1993) *Forensic Pathology*. Boca Raton, FL: CRC Press.

DiMaio VJM (1985) *Gunshot Wounds: Practical Aspects of Firearms, Ballistics, and Forensic Techniques*. Boca Raton, FL: CRC Press.

Gurdjian ES, Webster JE and Lissner HR (1950) The mechanism of skull fractures. *Radiology* **54**(3):313–338.

LaGarde L (1916) *Gunshot Injuries*. New York: William Wood.
Maples WR (1986) Trauma analysis by the forensic anthropologist. In: Reichs KJ (ed) *Forensic Osteology*, pp. 218–228. Springfield, IL: CC Thomas.
Merbs CF (1989) Trauma. In: İşcan MY and Kennedy KA (eds) *Reconstruction of Life from the Skeleton*, pp. 161–189. New York: Wiley-Liss.
Quatrehomme G and İşcan MY (1997) Beveling in exit gunshot wounds in bones. *Forensic Science International* **89**:93–101.
Quatrehomme G and İşcan MY (1997) Postmortem skeletal lesions. *Forensic Science International* **89**(3):515–565.
Quatrehomme G and İşcan MY (1998) Analysis of beveling in gunshot entrance wounds. *Forensic Science International* **93**:45–60.
Quatrehomme G and İşcan MY (1998) Gunshot wounds to the skull: comparison of entries and exits. *Forensic Science International* **94**:141–146.
Quatrehomme G and İşcan MY (1999) Characteristics of gunshot wounds in the skull. *Journal of Forensic Science* **44**:568–576.
Reichs KJ (1998) *Forensic Osteology. Advances in the Identification of Human Remains*, 2nd edn. Springfield, IL: Charles C Thomas.
Simonin C (1962) *Medecine Legale Judiciaire*. Paris: Librairie Maloine.
Spitz WU (1980) Gunshot wounds. In: Spitz W and Fisher R (eds) *Medicolegal Investigation of Death*, pp. 216–274. Springfield, IL: Charles C Thomas.
Ubelaker DH and Adams BJ (1995) Taphonomic indicators. *Journal Forensic Science* **40**: 509–512.

Stature Estimation from the Skeleton

T Sjøvold, Stockholm University, Stockholm, Sweden

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0770

Introduction

Stature provides one aspect of an individual's physiognomy and one piece of information that may be an aid in individual identification. In situations where the corpse is severely mutilated, decomposed or represented by skeletal remains only, the stature of the individual may be estimated by means of measurements of the skeleton, in certain cases after necessary maceration at a forensic laboratory. Such estimation is based on the relations between skeletal elements and stature. As a rule of thumb, the larger the skeletal element, the taller the individual. This means that, theoretically, any measurement of any bone or measurements of combinations of bones of an individual reflect that individual's stature. Because of individual variation, however, a stature estimate is always hampered with a probable error, although each method of estimating stature from skeletal measurements aims at estimating the stature as exactly as possible.

Most methods of estimating stature from the skeleton are based on the long bones of the upper and lower extremities. Others deal with stature estimation from parts of the skeleton, either single bones or combinations of bones. Yet others estimate stature from incomplete bones, normally parts of long bones.

Bones recovered in forensic situations may either have been lying on the surface, in the ground or in water. The six long bones of the skeleton are generally composed of a shaft forming a tube of compact bone with a marrow cavity along the middle of the tube, whereas the interior of the bone ends has a rigid, spongy structure, covered by a thin layer of compact bone. Even the compact bone consists of a complicated, microscopic system with a number of pores for small vessels for blood supply and transportation. Therefore, if not completely fresh, a bone is generally more or less moist when recovered, particularly if it has been exhumed from the ground or from water. If it has not been lying exposed for a sufficiently long period of time so that it has become dry, it may tend to break, although this also depends on the amount of organic matter left in the bone. Even compact bone may tend to break, and the thin, compact bony layer at the bone ends may easily be damaged if not handled with care. It is self-evident that an undamaged bone facilitates further investigation. Every bone should be handled carefully when removed from a forensic site, to ensure that it is as complete as possible for measurements specified for stature estimation to be taken.

Considerations regarding age, sex and ethnicity may be made when estimating stature from the skeleton. Ethnicity may be understood either in terms of Caucasoid, Mongoloid or African origin or descent, or as inhabiting a particular country, or both. An argument for this is that an individual's stature is a result of both genetic and environmental (including nutritional) factors in a way which is not fully understood. Differences in skeletal body proportions in relation to stature may depend on each of these factors in combination.

Stature increases during childhood and through puberty, until all bone growth has ceased after adulthood is reached. Growth during this period is variable, and there is an individually more or less accentuated 'growth spurt'. Before estimating stature, it is therefore necessary first to determine if the

individual is an adult from the point of view of skeletal growth. As for different ethnic groups, body proportions may vary because of selective adaptation to different kinds of climatic zones characteristic of each group.

Some words of caution should be given when estimating stature from the skeleton or judging the quality of a given method. Humans of the same population vary in body proportions, even individuals known to have the same stature. This means that for every given stature there are individuals with long trunks and short extremities or short trunks and long extremities, although the proportions are centered around mean population values. Because of this, for every estimate of stature from skeletal measurements there is an uncertainty because of the variability within the population. In general, the higher the correlation between the skeletal measurement(s) and the stature, the more accurate an estimate of the stature may be. In particular, this applies to the long bones or combinations of long bones. Conversely, if the part of the skeleton measured is small, greater uncertainty may be expected because of the lower correlation between stature and skeletal measurement. If the whole skeleton can be measured, or at least measurements of the single bones adding up to the skeletal height, allowing for the missing soft tissue, this may be preferable. Indeed, stature calculated from skeletal height has been used to represent stature when developing formulas for estimating stature from long bones (see Methods below).

In some cases methods have been tested in actual forensic cases when positive identifications have been established by means of other evidence. If the deviation between the estimate and the known stature is found to be small, this is taken to be a good property of the method. However, this may not mean that such a method is bad if there are larger deviations. It should be borne in mind that a particular method aims at a good result on average, which may not necessarily mean an optimum result will be achieved in each individual case. For every method of estimating stature from the skeleton, there will be a smaller or greater difference between estimated stature and actual stature. Ideally, this difference should be zero; that is, that the estimate of the stature is exact. What is sought therefore is a method where bias is as small as possible.

Stature as a Concept

Stature is not a fixed value for any individual at any age but is influenced by different factors. Every individual's stature tends to decrease during the period from getting up to going to bed. This decrease is due to the elasticity and compression of intervertebral disks and joint cartilage and load carried by the body during walking upright or sitting. Such a load may be due to both the body weight and the actual loads carried or lifted. Extreme reduction of stature as a result of carrying of heavy loads have been reported – up to some 10 cm in some cases – although a decrease of 1–2 cm may be regarded as normal. The time lapse or repeatability of exposure to the load is an important factor in the decrease, as is the period of rest for the body to restore the amount of vertical elasticity. The possible decrease in stature may also be related to stature itself: during the period out of bed, a tall and obese individual may have greater potential for stature decrease than a short and slender individual.

Another important factor is age, because the elasticity of the intervertebral disks and cartilage decrease by age. With increasing age there is a general tendency towards stature decrease. The decrease is generally regarded to be approximately 6 mm per decade after the age of 30. When estimating stature, the age factor may be accounted for by first calculating the maximum stature attained by the individual and then adjusting for decrease due to age; however, stature decrease due to age may not be completely regular, and individual variation occurs. For instance, in old age stature decrease may be substantial during a short period of time, owing to changes in posture or gait that may affect some individuals more than others. This means that information of stature given in, for example, passports, driving licences or military records are at most fair approximations of a feature which is not exactly fixed. Sometimes, recorded stature may not even have been measured but entered on the basis of oral information. Nevertheless, stature or estimates of stature provide important information for identification purposes.

Materials for Stature Estimation Methods

All methods of estimation of stature from the skeleton are based on different kinds of samples from the populations they are regarded as representing, or from combinations of samples from different populations. Knowledge of some sort of stature and bone length is always required in order to develop any method for stature estimation, although in most connections direct information is only known for one of the two kinds of data, either stature or bone length. The main effect of this is that, if one kind of data is not known but has to be estimated, the correlations between stature and bone length tend to be underestimated, leading to a slightly higher standard error of stature estimates in comparison to methods based on both known stature and known bone lengths. The

idea of stature estimation is, however, to provide an estimate of the maximum stature attained by an individual, which may be adjusted depending on information about the age of missing persons from whom the skeletal remains may derive. Five different kinds of source materials are encountered among existing methods for stature estimation:

1. In the ideal case, both (maximum) stature when living and bone lengths after death. Only a few cases exist where this condition is satisfied, mostly related to victims and repatriation of identified individuals temporarily buried during the World War II, although there are other cases where for some reason both stature during life and bone lengths after death have been recorded. Reported forensic cases where positive identification has been successful, where documented stature and bone measurements exist, provide additional data for a databank for further development of the methodology. For this kind of method both stature and bone lengths are primary data.
2. Dissection-room materials, that is cadavers and corresponding macerated bones measured. In this case, cadaver length has to be converted to stature by subtraction of postmortem extension of the length of the cadaver. Cadavers are known to *increase* in length after rigor mortis, due to the loosening of the intervertebral disks and slackening of the vertebral curvature. This increase is considered to be of the magnitude of 2–2.5 cm. In some cases the cadavers have been mounted, restoring the spinal curvature in order to approximate standing posture when living, and in other cases the lengths of cadavers have been measured on a dissection-room table. Although in the first case the stature of the mounted cadaver has been regarded as a measurement of the individual's stature, age effects may not be completely ruled out. However, as for cases where cadavers have been measured in a supine position, the mean age of individuals making up such samples is generally fairly high, and it may be argued that reduction in stature due to age may to a large degree be compensated by the postmortem extension of cadaver length. For this kind of method stature is not primary data, whereas bone lengths are.
3. Somatometric materials, that is stature of individuals have been measured and their physical extremity proportions measured. The extremity proportions are converted to bone lengths by means of mathematical conversion formulas. For this kind of method stature is primary data, whereas bone lengths are not.
4. Somatometric materials using X-ray, that is living individuals measured and their extremities X-rayed. This is a variation of (3), as measurements of the bones on the radiographs are used. The method has the drawback that the X-ray beam must be positioned precisely in relation to the bone for each individual in order to produce the measurement aimed at, a problem that may create additional, random variation. Using X-ray methods, the degree of magnification of bone lengths is a function of the distance of the X-ray film to the bone and the X-ray source. The X-ray beam also has to be positioned in exactly the same way for every new application. For this kind of method stature is primary data, whereas bone lengths are not.
5. Methods where stature has been estimated from skeletal height, and bone measurements have been taken from the same skeleton. For this kind of method bone lengths are primary data, whereas stature is strictly speaking not.

Methods for Stature Estimation

The methods in use may be divided into four categories based on:

- least squares regression equations;
- other regression principles;
- stature:bone length ratios;
- skeleton height and adjustment for missing soft tissue.

For each of these methods sex and race may also be a matter of concern. These two factors become particularly important when sex or race or both are unknown, a situation which may frequently be encountered in forensic cases, particularly when only skeletal remains have been recovered. The ultimate aim, however, is to be able to produce an estimate of the stature which is as close to the actual stature of the unknown individual as possible.

A large number of studies on stature estimation have been published. Most of these concern stature estimation based on measurements of the long bones of the extremities. An important reason for that is that such methods minimize the error caused by biological variability. Although methods have also been developed using other bones, the uncertainty tends to grow so that the accuracy becomes too unsatisfactory. The same concerns stature estimation of children, where methods have been developed based on immature skeletal parts. Even in this case there are large uncertainties that depend on the rate of skeletal development of the children. Therefore these aspects are not discussed further here, although the general considerations made are also applicable in

such cases, and the reader is referred to the Further Reading list for more information. The only method discussed that is not based on measurements of the long extremity bones concerns stature estimation based on skeletal height.

Stature estimation from long bones based on least squares regression

Most methods for estimating stature from long bones are based on this principle. It implies that the standard error is as small as possible from a mathematical point of view. The method is based on the principle that it minimizes the sum of squared deviations from the individual stature:bone length values which represent the sample used for calculating the formula for estimating the stature. This kind of method has both advantages and disadvantages. An advantage is that the errors are minimized when regarding a large number of individuals from the population on which the method is based. On the other hand, a disadvantage is that these methods, strictly speakng, should not be used for individuals from another population, and the longest bone does not reproduce the tallest individual nor the shortest bone the shortest individual, but tend to underestimate the tallest and overestimate the shortest. Experience has also shown that secular trends in a population may cause problems because of alterations of stature:bone length proportions in the population, so that measurements of individuals of a later generation may become incompatible with the proportions on which the original method was based. Significant changes of this nature are known to have taken place, for example in the USA between World War II and the Korean War.

The term 'population' in this connection is, however, not a true biological concept. The only inference that may be made is that a sample used for any method for stature estimation consists of humans displaying certain relations between stature and bone measurements, which may be given in terms of sample size, mean values, standard deviations and correlations.

Another way of looking at the deviation between estimated and actual is to regard the confidence contours of the regression line slope. This takes into account that each method is based on a sample, from which the regression line is calculated. However, since sample bias may exist, the slope of the regression line is also determined with some uncertainty, and may be allowed to vary, that is be less steep or steeper than the slope calculated from the regression line. Since the regression line passes through the mean value of the sample, estimates close to the mean value are connected with the smallest probable errors. The further away from the mean value, the larger the probable error. The reason is that every estimate is related to the population from which the sample used for the method derives, which means that every new case applied to a regression equation is formally regarded as belonging to the same population.

Least squares regression may also be presented in terms of multiple regression, by means of which stature may be estimated based on several bone lengths, taking the correlation between the bones into consideration. When the most efficient bones are incorporated into a multiple regression formula, the standard error of estimate may be further reduced. The number of bones successfully incorporated in a regression formula in order to estimate stature has, however, a limit. This limit is determined by the amount of information about the stature which is provided by the bone measurements. If still more bones are added, the standard error tends to increase compared with its minimum value, because more variability than information is introduced. In most cases, this limit is reached by means of two or three bone measurements.

Because the mean stature of males is higher than that of females in every population, regression equations have to be developed for both males and females. Different regression equations for males and females are also needed because the body proportions of males and females differ. These differences are partly due to differences in mean stature, but may also be due to female skeletal adaptations of the pelvic girdle in order to be able to give birth, broadening the pelvic inlet, and adapting the inclination of the shaft of the femur to the broadened pelvis.

Most methods for stature estimation based on skeletal parts other than the long extremity bones are based on least squares regression (see Further Reading).

Stature estimation based on other regression principles

Since regression tends to overestimate the stature of the shortest individuals and underestimate the stature of the tallest individuals, an alternative has been to use a so-called regression II model, also known as the reduced major axis. The line calculated is related to least squares regression, and may be denoted least triangles regression. It minimizes the sum of triangles formed by looking at the deviation from the line for all individual stature:bone length values in the sample, with one side formed by the deviation from the line along the stature axis and the other at a right angle along the bone length axis, with the line itself forming the third side of the triangle. Theoretically, using this method, the longest bone reproduces the tallest stature and the shortest bone reproduces the

shortest stature.

A generalization of this method based on information on mean stature and mean long bone lengths from worldwide populations is called the weighted line of organic correlation because differences in sample size are taken into consideration, so that a large sample is given more weight than a small sample. This kind of method has even been shown to be independent of sex because differences in stature: bone length proportions between males and females as expressed by sample mean values appear to be primarily related to stature differences; a similar argument may also be applied with regard to ethnicity. The changes in proportions among females because of skeletal adaptation for giving birth appear to be negligible in comparison. Because of the regularity of the stature and long bone mean values for random, worldwide populations, the method also aims to cover changes in stature of a population over time due to secular trends, and may also be used if race and ethnicity are unknown. On the other hand it is clear that, if ethnicity is known, a method based on that ethnic group will be better than a method based on worldwide populations.

Stature estimation based on stature: bone length ratios

Another worldwide survey of the relationship between mean stature and mean femur lengths has shown that this relationship appears to be rather stable throughout the world. Though presented as a femur:stature ratio, it is actually the stature:femur ratio that is used for stature estimation of an unknown individual. The stature:femur ratio is derived as 3.74, corresponding to a femur:stature ratio of 26.75, based on maximum length of femur (see Measurements below). Therefore stature may be estimated using this ratio instead of using more complicated formulas when the femur is available. It is argued in favor of such a method that every regression method has an error, so using a simple ratio instead will most likely provide a stature estimate that is within the standard error of a regression formula. Tests of this principle on skeletons of identified war casualities have shown it to do better than least squares regression; that is, estimates with less deviation from the known stature for tall individuals than the corresponding least squares regression equation.

Although in the beginning it was assumed that such a ratio would be independent of sex and ethnicity, tests have shown that this is not strictly the case, and different ratios may be applied depending on sex. As for differences between ethnic groups, tests have furthermore shown that when the ethnicity is not known, which is frequently the case when only skeletal remains of an individual are recovered, the error made when using a generalized ratio is less than when using a ratio for a certain ethnic group if it later turns out that the individual belonged to another ethnic group than the one on which the ratio was based. So far, however, the standard error of this ratio has not been derived, though it may be expected that it is larger than that of the reduced major axis. In comparison to regression and the weighted line of organic correlation, use of the stature:femur ratio will rather tend to overestimate the stature of the tallest individuals and underestimate the stature of the shortest. Furthermore, the stature:femur ratio represents a line that passes through the origin, whereas both regression and the weighted line of organic correlation are lines adapted only to the data, and is not forced to pass through the origin as well. In the last major study by the group dealing with the ratio, however, it was found that a generic regression based on worldwide data, almost identical to the corresponding equation based on the weighted line of organic correlation, performed even better than the ratio, meaning that mean body proportions change when stature changes. This fact was even noted by the Frenchman Leon Manouvrier in 1892, who was the first to publish a method on stature estimation from the skeleton for general use, although it is now only of historical interest.

Stature estimation based on skeletal height

Regarding the skeletal elements adding to stature, it may be claimed that, standing upright, stature comprises the height of the skull, the height of the vertebral column, the sacrum to the level of the top of the femoral head, the length of the femur in its natural position, the physiological length of the tibia, and the height of the articulated talus and calcaneus from the foot, to which is added the thickness of soft tissue.

The height of the articulated calcaneus and talus of the foot is measured as in the standing position, the physiological length of the tibia is measured from the distal articular surface to the proximal articular surface, and the length of the femur in the natural (bicondylar) position is measured as defined under Measurements below. Since it is difficult to reconstruct the curvature of the vertebral column from a disarticulated skeleton, the height of the anterior (ventral) side of all vertebral bodies, from the first sacral vertebra to the second cervical vertebra, including the odontoid process (dens axis), may be regarded as the length of the vertebral column. The first cervical vertebra (atlas) is excluded because ontogenetically the body of the atlas is assimilated with that of axis to form the odontoid process of the axis. The height of

the skull is measured as the distance between the basion and the bregma. This measurement is regarded as the height of the skull because it is easy to measure and because there are individual variations as to how to define the height of the skull when regarding a living individual. To the accumulated sum of measurement is added 10 cm if the sum is less 153.5 cm, 10.5 cm if it is between 153.5 cm and 163.4 cm and 11.0 cm if it is 163.5 cm or more, to compensate for the net thickness of missing soft tissue.

Stature estimation formulas based on least squares regression have even been developed for parts of the elements included in the skeletal height, although the methods make the application population-specific, and they are therefore not discussed further here.

Measurements

Apart from the measurements described in connection with the estimation of stature from skeletal height, all measurements should be made with an osteometric board. The measurements described are those normally used for the six long bones.

- *Humerus*, maximum length. The distance from the medial margin of the trochlea to the highest point of the humeral head. The humeral head is placed against the vertical wall of the osteometric board and the block at the medial margin of the trochlea. The bone is moved in any direction until maximum length is obtained.
- *Radius*, maximum length. The greatest distance from the tip of the styloid process to the margin of the radial head, measured by the same procedure as for the humerus.
- *Ulna*, maximum length. The maximum distance between the highest point of the olecranon and the most distal point of the styloid process, measured by the same procedure as for the humerus.
- $\mathit{Femur}_{\mathrm{max}}$, maximum length. The distance between the highest point of the femoral head and the most distal point of the medial condyle while the bone is lying on the osteometric board with the medial condyle touching the vertical wall. The bone is rotated until maximum length is obtained.
- $\mathit{Femur}_{\mathrm{phys}}$, physiological length, length in the natural position (bicondylar length). The vertical distance from the femoral head to the vertical wall of the osteometric board when both condyles are placed against the vertical wall.
- *Tibia*, total length. The distance from the tip of the medial malleolus to the lateral part of the lateral condyle. The tibia is placed with the dorsal side on the osteometric board with the apex of the malleolus against the vertical wall, the longitudinal axis of the bone at right angles to the vertical wall, and the block placed against the lateral part of the lateral condyle. The measurement should not be confused with the maximum length, which includes the intercondylar eminences.
- $\mathit{Tibia}_{\mathrm{phys}}$, physiological length. The distance from the center of the proximal, medial articular surface and the base of the medial malleolus, at the articulation with the talus. Measured with a large spreading caliper.
- *Fibula*, maximum length. The direct distance between the most proximal and the most distal points of the fibula. Note, however, that when drying, the fibula may tend to bend to a curve, in which case this measurement is no longer possible.

In the German literature, following the measuring system of Rudolf Martin, all measurements are measurement no. 1 for each particular bone, except for the bicondylar length of femur and the physiological length of tibia, which are measurements no. 2. In Appendix 1, the latter measurements are indicated by the subscript ‘phys’.

See also: **Anthropology**: Morphological Age Estimation; Sex Determination; Determination of Racial Affinity; Excavation and Retrieval of Forensic Remains.

Further Reading

Feldesman MR and Fountain RL (1996) ‘Race’ specificity and the femur/stature ratio. *American Journal of Physical Anthropology* **100**:207–224.

Fully G (1956) Une nouvelle méthode de détermination de la taille. *Annales de Médicine Légale* **36**:266–273.

Hrdlička A (1939) *Practical Anthropometry*. Philadelphia: Wistar.

Jantz RL, Hunt DR and Meadows L (1995) The measure and mismeasure of the tibia: implications for stature estimation. *Journal of Forensic Sciences* **40**:758–761.

Krogman WM and İşcan MY (1986) *The Human Skeleton in Forensic Medicine*. Springfield, IL: Thomas.

Lundy JK and Feldesman MR (1987) Revised equations for estimating living stature from the long bones of the South African Negro. *South African Journal of Science* **83**:54–55.

Manouvrier L (1893) La détermination de la taille d’après les grands os des membres. *Mémoires de la Société d’Anthropologie de Paris* **4**:347–402.

Olivier G, Aaron C, Fully G and Tissier G (1978) New estimation of stature and cranial capacity in modern man. *Journal of Human Evolution* **7**:513–518.

Pearson K (1899) On the reconstruction of stature of prehistoric races. *Transactions of the Royal Society, Series A* **19**:169–244.

Rösing FW (1988) Körperhöhenrekonstruktion aus

Skelettmaßen. In: Martin R and Knußmann R (eds) *Anthropologie. Handbuch der vergleichenden Biologie des Menschen. Begründet von Rudolf Martin*, vol. 1, pp. 586–600. Stuttgart: Fischer.
Sjøvold T (1990) Estimation of stature from long bones utilizing the line of organic correlation. *Human Evolution* **5**:431–447.
Stevenson PH (1929) On racial differences in stature long bone regression formulae, with special reference to stature reconstruction formulae for the Chinese. *Biometrika* **21**:303–321.
Trotter M (1970) Estimation of stature from intact limb bones. In: Stewart TD (ed.) *Personal Identification in Mass Disasters*, pp. 71–84. Washington, DC: National Museum of Natural History.
Trotter M and Gleser GC (1951) The effect of ageing on stature. *American Journal of Physical Anthropology, New Series* **9**:311–324.
Trotter M and Gleser GC (1952) Estimation of stature from the long bones of American Whites and Negroes. *American Journal of Physical Anthropology, New Series* **10**:463–514.
Trotter M and Gleser GC (1958) A re-evaluation of stature based on measurements taken during life and long bones after death. *American Journal of Physical Anthropology, New Series* **16**:79–124.
Trotter M and Gleser GC (1977) Corrigenda: Estimation of stature from long limb bones of American Whites and Negroes. *American Journal of Physical Anthropology* **47**:355–356.
Xiang-Quing S (1989) Estimation of stature from intact long bones of Chinese males. *Canadian Society of Forensic Sciences Journal* **22**:123–126.

Appendix 1

Tables for stature estimation

The following tables provide some of the equations for stature estimation based on long bones, for Caucasian, Afro-Americans, Mongoloid or Asian populations, based on least squares regression and the weighted line of organic correlation. All estimates relate to the maximum stature of adult individuals. Subsequent adjustment (reduction) of stature may be carried out in order to match individuals over 30 years old. The use of long bones to estimate stature is recommended in preference to other methods because of the smaller standard error, with the exception of the method based on skeletal height described in the text. Some of the methods are based on long bones from one side of the body only, as stated in the tables. If bones from the other side are used, long bones of the right side are generally 1–2 mm longer than the left side in the upper extremities and 1–2 mm shorter in the lower extremities because of skeletal asymmetry. All measurements are in centimeters.

Least squares regression equations

Caucasian

American caucasian males (adapted from Trotter and Gleser, 1958, Table 12; mean stature 174.6 cm)

Stature = 2.89 humerus	+78.10	± 4.57
Stature = 3.79 radius	+ 79.42	± 4.66
Stature = 3.76 ulna	+ 75.55	± 4.72
Stature = 2.32 femur	+ 65.53	± 3.94
Stature = 2.60 fibula	+ 75.50	± 3.86
Stature = 1.82 (humerus + radius)	+ 67.97	± 4.31
Stature = 1.78 (humerus + ulna)	+ 66.98	± 4.37
Stature = 1.31 (femur + fibula)	+ 63.05	± 3.62

Formulas including tibia are omitted because of ambiguity as to how the bone has been measured (see Jantz RL *et al.*, 1995).

American caucasian females (adapted from Trotter and Gleser, 1952, Table 13; mean stature 160.7 cm)

Stature = 3.36 humerus	+ 57.97	± 4.45
Stature = 4.74 radius	+ 54.93	± 4.24
Stature = 4.27 ulna	+ 57.76	± 4.30
Stature = 2.47 femur	+ 54.10	± 3.72
Stature = 2.93 fibula	+ 59.61	± 3.57

Formulas including tibia are omitted because of ambiguity as to how the bone has been measured (see Jantz RL *et al.*, 1995).

European males (adapted from Olivier *et al.*, 1978, Table 3; left side, mean stature 170.4 cm)

Stature = 3.1906 humerus	+ 64.19	± 4.03
Stature = 3.9582 ulna	+ 66.71	± 4.47
Stature = 2.4202 $femur_{phys}$	+ 58.33	± 3.48
Stature = 2.6841 fibula	+ 70.96	± 3.44
Stature = 2.257 humerus + 1.586 ulna	+ 53.29	± 3.79
Stature = 1.121 humerus + 1.760 $femur_{phys}$	+ 51.56	± 3.34
Stature = 1.234 ulna + 1.935 $femur_{phys}$	+ 48.41	± 3.34
Stature = 0.781 ulna + 2.492 fibula	+ 66.43	± 3.37
Stature = 1.213 $femur_{phys}$ + 1.548 fibula	+ 56.93	± 3.10
Stature = 0.593 humerus + 0.983 $femur_{phys}$ + 1.384 fibula	+ 53.90	± 3.05

Formulas involving radius and tibia are excluded because of deviating measurements.

European females (adapted from Olivier *et al.*,

1978, Table 4; left side, mean stature 160.2 cm)

Stature = 3.0882 humerus	+ 62.31	± 3.62
Stature = 4.0931 ulna	+ 63.71	± 3.53
Stature = 2.0960 $femur_{phys}$	+ 70.20	± 3.56
Stature = 1.549 humerus + 2.408 ulna	+ 54.38	± 3.23
Stature = 1.661 humerus + 1.240 $femur_{phys}$	+ 54.33	± 3.13
Stature = 2.468 ulna + 1.236 $femur_{phys}$	+ 48.96	± 2.99

Formulas involving radius and tibia are excluded because of deviating measurements.

Afro-American and African

Afro-American males (adapted from Trotter and Gleser, 1958, Table 12; mean stature 173.8 cm)

Stature = 2.88 humerus	+ 75.48	± 4.23
Stature = 3.32 radius	+ 85.43	± 4.57
Stature = 3.20 ulna	+ 82.77	± 4.74
Stature = 2.10 femur	+ 72.22	± 3.91
Stature = 2.34 fibula	+ 80.07	± 4.02
Stature = 1.66 (humerus + radius)	+ 73.08	± 4.18
Stature = 1.65 (humerus + ulna)	+ 70.67	± 4.23
Stature = 1.20 (femur + fibula)	+ 67.77	± 3.63

Formulas including tibia are omitted because of ambiguity as to how the bone has been measured (see Jantz RL, *et al.*, 1995).

South African males (skeletal height) (adapted from Lundy and Feldesman, 1987, Table 2; mean stature not given)

Stature = 2.899 humerus	+ 60.212	± 3.834
Stature = 3.196 radius	+ 72.139	± 3.643
Stature = 2.961 ulna	+ 72.700	± 3.727
Stature = 2.403 $femur_{phys}$	+ 45.721	± 2.777
Stature = 2.427 $tibia_{phys}$	+ 60.789	± 2.780
Stature = 2.515 fibula	+ 58.999	± 2.980
Stature = 1.288 ($femur_{phys}$ + $tibia_{phys}$)	+ 46.543	± 2.371

For skeletal heights under 153.5 cm, add 10.0 cm; for skeletal heights between 153.5 cm and 163.4 cm, add 10.5 cm; for skeletal heights of 163.5 cm or above, add 11.0 cm to estimate stature.

Afro-American females (adapted from Trotter and Gleser, 1952, Table 13; mean Stature 160.9 cm)

Stature = 3.08 humerus	+ 64.67	± 4.25
Stature = 3.67 radius	+ 71.79	± 4.59[a]
Stature = 3.31 ulna	+ 75.38	± 4.83
Stature = 2.28 femur	+ 59.76	± 3.41
Stature = 2.49 fibula	+ 70.90	± 3.80

[a] Modified according to Trotter and Gleser 1977.

Formulas including tibia are omitted because of ambiguity as to how the bone has been measured (see Jantz RL *et al.*, 1995).

South African females (skeletal height) (adapted from Lundy and Feldesman, 1987, Table 1; mean stature not given)

Stature = 3.291 humerus	+ 45.893	± 3.715
Stature = 4.161 radius	+ 47.120	± 3.387
Stature = 3.827 ulna	+ 47.574	± 3.629
Stature = 2.769 $femur_{phys}$	+ 27.424	± 2.789
Stature = 2.485 $tibia_{phys}$	+ 55.968	± 3.056
Stature = 2.761 fibula	+ 47.574	± 3.168
Stature = 1.410 ($femur_{phys}$+ $tibia_{phys}$)	+ 34.617	± 2.497

For skeletal heights under 153.5 cm, add 10.0 cm; for skeletal heights between 153.5 cm and 163.4 cm, add 10.5 cm; for skeletal heights of 163.5 cm or above, add 11.0 cm to estimate stature.

Asian and Mongoloid

Mongoloid males (adapted from Trotter and Gleser, 1958, Table 12; mean stature 167.5 cm)

Stature = 2.68 humerus	+ 83.19	± 4.25
Stature = 3.54 radius	+ 82.00	± 4.60
Stature = 3.48 ulna	+ 77.45	± 4.66
Stature = 2.15 femur	+ 72.57	± 3.80
Stature = 2.40 fibula	+ 80.56	± 3.24
Stature = 1.67 (humerus + radius)	+ 74.83	± 4.16
Stature = 1.68 (humerus + ulna)	+ 71.18	± 4.14
Stature = 1.22 (femur + fibula)	+ 63.05	± 3.18

Formulas including tibia are omitted because of ambiguity as to how the bone has been measured (see Jantz RL *et al.*, 1995).

Chinese males (adapted from Xiang-Qing, 1989, Tables 3–4; mean stature 164.3 cm

Stature = 2.66 humerus	+ 82.64	± 4.13
Stature = 3.49 radius	+ 82.71	± 4.14
Stature = 2.86 ulna	+ 92.82	± 4.47
Stature = 2.30 femur	+ 64.36	± 3.48
Stature = 2.22 tibia	+ 85.34	± 3.87
Stature = 2.54 fibula	+ 76.15	± 3.81
Stature = 1.47 (humerus + radius)	+ 84.78	± 4.31
Stature = 1.57 (humerus + ulna)	+ 76.82	± 4.12
Stature = 1.10 (femur + tibia)	+ 77.78	± 3.91
Stature = 1.25 (femur + fibula)	+ 66.95	± 3.72

The formulas are based on the age group 21–30 years.

Weighted line of organic correlation

Caucasian, both sexes (adapted from Sjøvold, 1990, Table 1; all equations based on weighted mean values of 9–23 samples representing 2308–8577 individuals depending on bone)

Stature = 4.74 humerus	+ 15.26	± 4.94
Stature = 4.03 radius	+ 69.96	± 4.98
Stature = 4.65 ulna	+ 47.96	± 4.96
Stature = 2.63 femur	+ 49.96	± 4.52
Stature = 3.10 $femur_{phys}$	+ 28.82	± 3.85
Stature = 3.02 tibia	+ 58.94	± 4.11
Stature = 3.78 fibula	+ 30.15	± 4.06

All ethnic groups or, if ethnicity is unknown, both sexes (adapted from Sjøvold, 1990, Table 2; all equations based on weighted mean values of 22–44 samples representing 3232–10 573 individuals, depending on bone)

Stature = 4.62 humerus	+ 19.00	± 4.89
Stature = 3.78 radius	+ 74.70	± 5.01
Stature = 4.61 ulna	+ 46.83	± 4.97
Stature = 2.71 femur	+ 45.86	± 4.49
Stature = 3.01 $femur_{phys}$	+ 32.52	± 3.96
Stature = 3.29 tibia	+ 47.34	± 4.15
Stature = 3.59 fibula	+ 36.31	± 4.10

ANTHROPOMETRY

M Y İşcan, Adli Tıp Enstitusu, Istanbul Univesitesi, Istanbul, Turkey

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0416

Introduction

Two of the most commonly used methods in forensic anthropology are the metric (anthropometry) and morphological (anthroposcopy) assessment of the characteristics of living and skeletal remains. The origins of these methods go back to the eighteenth century, when there was a great deal of interest in human diversity, especially of those who had just been 'discovered' by the colonizing European nations.

In the following century the methods were further developed for the purpose of tracing the ethnic origins of human populations. In the first half of the twentieth century, these methods (especially anthropometry) played an important role in understanding human variation and evolution. In the second half of the century, as the concept of race fell from the research paradigm of anthropology, anthropometry and anthroposcopy took on a new role, especially in the fields of human engineering and forensic anthropology.

In essence, anthropometry is a systematic study of human measurements. Dimensions taken are internationally agreed and standardized linear and angular measurements. If the dimensions are taken from a particular region of the human body, the technique may have a different name, such as cephalometry for the head, craniometry for the skull, odontometry for the teeth and osteometry for the skeletal system. There are many different data collection forms that list measurements to be taken from a living person and from a skeleton (skull and the postcranial skeleton, respectively). To these may be added others to suit the research interests of the investigator. The measurements now form the very foundation of the methodology used to study human growth, differences between populations, and health and disease-related problems.

Anthroposcopy is more complex and requires years of experience in assessing the variation of a feature in different populations. The assessment criteria used in this technique are also at the very root of anthropometry; that is, anything that has a morphological variation can be observed with the naked eye and can also be measured. Many anthropologists therefore rely heavily on first-hand observation of how large or small, short or long, rough or smooth, robust or gracile are the features being examined. Yet many observable characteristics cannot easily be quantified. These features should be recorded, as such, without

any quantification. There are also blank forms for recording those anthroposcopic features that have traditionally been observed in both the living person and the skeleton. To these one must add many observable features of dentition.

Photographic comparison has provided lists of anthropometric landmarks and dimensions that can be modified for use in comparing two different facial images. The same approach can be applied to estimate the age of a living person from a photograph. This is commonly done if the picture or an Internet-based graphic file is thought to be that of a naked child. In this case, secondary sexual characteristics, such as breast development and axillary and pubic hair in a girl, and facial hair in a boy, may have to be assessed morphologically rather than metrically.

These metric and morphological characteristics can be assessed in both the living and in skeletons. Their use, however, is more common in skeletal remains. This article reviews the use of anthropometric and anthroposcopic characteristics in the study of the forensic anthropology of skeletal remains.

Anthropometric Analysis of Skeletal Remains

Forensic anthropology has benefited considerably from both anthropometric and anthroposcopic techniques. Many biological differences in terms of sex, race and stature depend largely upon human size and shape. To assess population variation, a set of measurements can be obtained from skeletal collections of known demographic characteristics. Studies have indicated that determination of sex or other demographic characteristics require population-specific studies. A sample of 50 or more individuals per subgroup (e.g. 50 males of a black sample) may fulfill the minimum requirements of the study, assuming that there are that many skeletons in the collection. **Table 1** lists descriptive statistics of osteometric dimensions in males and females of different populations from dissection room collections in South Africa (Negroid and Caucasoid) and Japan (Mongoloid). The dimensions are from both the skull, including the mandible, and the long bones of the postcranial skeleton. The differences between the samples can be analyzed using Student's *t*-test or even analysis of variance statistics. Many dimensions show similarity between, for example, the Japanese and Whites, or between Whites and Blacks. It is clear that size plays an important role in population difference; however, most anthropometric analysis involves multiple dimensions. Anthropologists have relied heavily on this factor in determining sex and race and estimating stature. Anthropometric dimensions, as such, assume that any two or more groups can be analyzed in terms of their body size.

Sometimes differences in size between individuals, such as a male and a female, or populations may have to be assessed in terms of shape. Shape generally falls into anthroposcopic determination. For example, cranial (and cephalic) shapes alone can separate one group from another. Although this observation can be made with a naked eye, it is easier if the observation can be quantified. A traditional solution to assessing shape using anthropometry has been the creation of an index with two dimensions describing the structure. Therefore an index is the percent expression of the ratio (proportion) of a smaller dimension over a larger one; for example, the cranial index, which is calculated as 100 * cranial breadth/cranial length. **Table 2** lists several commonly used cranial indices for the same populations shown in Table 1. The advantage of an index is severalfold. It eliminates the size difference between individuals; that is, a tall person can be compared with a short one. Elimination of size difference also provides suggestions about the shape (morphology) of a structure. For example, a cranial index of 70.9 in Blacks describes the shape of that person as dolichocranic or longer skull (dolichocephalic if the dimensions are taken from a living person). The same index describes a Japanese skull as mesocranic, that is, rounder. Gnathic index is formed by the ratio of the basion–prosthion length divided by the basion–nasion length. The purpose of the index is to show which population has the most protruding chin. It is clear that Blacks have the largest value. The following is an example of how the cranial index has been classified:

Cranial length (or length–breadth) index	*Range*
Dolichocranic	*x*–74.9
Mesocranic	75.0–79.9
Brachycranic	80.0–*x*

Many indices exist and others can be created if the investigator deems it necessary. However, for population comparative purposes, only those that are used worldwide are reported in a study. It also reduces interobserver errors, unlike anthroposcopic assessment, which depends on the observer's knowledge of the population.

Because anthropometry deals with body dimensions, statistics play an important role: in fact, many earlier statistical techniques were developed using anthropometric dimensions. Statistical techniques that are commonly employed in forensic anthropometric studies include discriminant function and regression analyses.

Table 1 Comparative osteometric dimensions in Japanese, Whites, and Blacks[a]

Measurements	*Japanese*			*Whites*			*Blacks*		
	No. Males	*Mean*	*SD*	*No.*	*Mean*	*SD*	*No.*	*Mean*	*SD*
Males									
Cranial length	46	178.7	7.7	53	187.6	5.4	45	186.9	5.0
Cranial breadth	45	142.3	5.9	53	140.2	6.1	45	132.5	5.6
Minimum frontal breadth	46	95.3	5.2	53	97.4	4.1	45	97.1	4.0
Bizygomatic breadth	46	135.2	5.8	53	129.2	4.3	45	130.4	4.6
Basion–nasion	47	101.2	4.2	53	102.1	5.2	44	101.5	3.5
Basion–bregma	45	138.5	5.0	53	136.4	4.8	45	133.7	4.7
Basion–prosthion	47	94.6	4.9	47	95.5	5.2	45	104.4	6.8
Nasion–prosthion	46	69.9	4.7	48	71.3	3.9	45	68.4	4.6
Mastoid height	47	33.0	2.3	53	34.5	3.6	43	30.9	3.1
Nasal height	47	51.7	2.6	53	53.5	3.6	45	48.2	4.4
Nasal breadth	47	26.2	2.1	53	24.7	2.1	45	27.8	2.5
Bicondylar length	44	73.8	4.4	48	76.8	5.5	45	79.2	4.7
Bicondylar breadth	44	123.8	6.0	46	116.8	5.3	45	114.8	5.0
Bigonial breadth	44	102.3	5.4	48	99.4	5.8	44	97.1	5.6
Minimum ramus breadth	44	32.3	3.4	46	31.3	3.5	45	35.4	3.3
Scapular height	42	154.2	10.3	55	160.6	9.0	1	136.0	
Clavicle length	44	147.6	6.5	46	157.0	11.5	2	148.5	14.9
Humeral length	44	297.4	10.4	55	335.0	17.9	41	327.7	14.7
Humeral head diameter	44	44.1	1.8	55	49.0	3.2	43	43.7	2.2
Humeral epicondylar breadth	44	59.8	2.3	55	64.3	3.9	43	61.5	6.0
Radial length	39	223.4	8.6	53	249.8	15.8	1	250.0	
Ulnar length	39	239.7	8.6	54	266.3	17.7	1	276.0	
Femoral maximum length	44	417.4	16.5	56	469.7	28.0	43	463.6	21.8
Femoral head diameter	45	46.0	2.0	56	48.5	2.7	42	45.5	2.3
Femoral midshaft circumference	45	85.4	4.6	56	93.2	6.1	44	89.7	5.6
Femoral distal breadth	42	80.7	3.1	56	84.6	4.6	43	79.5	4.1
Tibial length	47	333.8	15.3	53	382.2	29.0	43	395.0	20.8
Tibial proximal breadth	45	73.4	2.8	56	79.1	4.9	44	75.1	3.6
Tibial distal breadth	46	45.4	2.4	45	46.9	2.6	43	46.4	7.4
Tibial circumference at nutrient foramen	46	91.9	5.0	56	98.0	6.6	40	97.8	6.9
Fibular length	47	333.8	14.0	56	378.1	24.8	43	385.5	20.1
Females									
Cranial length	35	173.4	7.5	53	179.0	5.9	45	178.2	6.3
Cranial breadth	34	137.6	6.3	53	137.6	4.6	45	129.9	6.9
Minimum frontal breadth	35	91.3	5.2	53	93.5	4.5	45	93.4	4.4
Bizygomatic breadth	33	128.7	5.5	53	121.9	3.5	45	121.5	5.1
Basion–nasion	35	97.8	4.5	53	96.7	4.2	44	95.2	4.0
Basion–bregma	34	133.0	5.8	53	130.6	5.2	45	126.4	6.7

Table 1 Continued

Measurements	*Japanese*			*Whites*			*Blacks*		
	No. Males	*Mean*	*SD*	*No.*	*Mean*	*SD*	*No.*	*Mean*	*SD*
Basion–prosthion	32	94.1	5.0	51	90.6	5.2	45	97.1	6.1
Nasion–prosthion	29	69.1	5.1	50	66.0	5.0	44	65.6	4.3
Mastoid height	35	29.0	3.0	53	31.5	4.1	45	26.3	3.0
Nasal height	35	50.1	2.9	53	49.8	2.2	45	47.2	2.5
Nasal breadth	35	26.0	2.1	53	23.0	2.0	45	26.7	2.0
Bicondylar length	29	70.4	12.0	47	72.7	5.3	45	74.7	5.1
Bicondylar breadth	29	117.2	6.3	49	111.4	5.8	45	108.5	5.6
Bigonial breadth	29	96.1	4.9	49	92.2	5.1	45	89.7	6.3
Minimum ramus breadth	29	31.3	3.2	49	28.6	2.7	45	32.9	3.1
Scapular height	34	137.3	12.4	48	143.9	8.3			
Clavicle length	35	134.0	8.1	42	143.9	6.4			
Humeral length	35	276.9	17.1	48	309.4	16.7	45	295.0	14.8
Humeral head diameter	35	39.1	2.6	49	43.1	2.5	45	37.7	2.1
Humeral epicondylar breadth	35	52.1	3.8	48	55.9	2.8	45	53.4	3.6
Radial length	32	199.9	13.9	50	224.6	13.0			
Ulnar length	32	217.1	13.6	49	240.7	13.5			
Femoral maximum length	34	390.3	23.7	50	437.6	20.7	45	423.7	22.3
Femoral head diameter	34	40.7	2.5	50	43.0	2.4	44	39.2	2.6
Femoral midshaft circumference	34	78.1	6.3	49	84.7	5.5	45	78.8	4.7
Femoral distal breadth	34	72.2	4.3	49	75.1	3.3	45	69.5	4.8
Tibial length	34	310.1	20.8	50	359.5	18.8	44	357.5	20.5
Tibial proximal breadth	34	65.8	4.7	50	69.8	3.7	45	64.9	7.3
Tibial distal breadth	34	40.5	2.6	48	42.0	2.5	45	39.8	2.2
Tibial circumference at nutrient foremen	34	82.4	8.3	50	87.5	6.0	45	84.9	5.3
Fibular length	33	310.2	21.7	49	348.7	18.6	44	348.5	20.8

[a] Japanese sample is from Jikei University collection and White and Black samples from the Dart and Pretoria collections of South Africa.

Table 2 Comparative osteometric indices in Japanese, Whites, and Blacks

Indices	*Japanese*			*Whites*			*Blacks*		
	No.	*Mean*	*SD*	*No.*	*Mean*	*SD*	*No.*	*Mean*	*SD*
Male									
Cranial	45	79.8	4.1	53	74.8	3.6	45	70.9	3.1
Height–length	44	77.7	4.2	53	72.7	2.7	45	71.6	2.9
Height–breadth	44	97.5	4.5	53	97.5	5.5	45	101.1	5.2
Nasal	47	50.6	4.6	53	46.4	5.0	45	58.5	12.7
Gnathic	47	93.5	3.9	47	93.2	3.4	44	102.9	6.6
Female									
Cranial	34	79.5	4.6	53	76.9	3.3	45	72.9	3.8
Height–length	34	76.9	3.9	53	73.0	3.2	45	70.9	3.1
Height–breadth	34	96.8	4.5	53	95.0	4.3	45	97.5	5.8
Nasal	35	52.0	3.9	53	46.3	3.9	45	56.7	4.6
Gnathic	32	96.1	4.8	51	93.7	3.8	44	101.9	5.1

Anthroposcopic Analysis of Skeletal Remains

Anthroposcopic (or morphological) analysis has contributed considerably to the study of forensic anthropology. Comparison of photographic images with those of videotape can best be done by examining all facial features for morphological variation; published studies include tables of those features that play a role in assessing the similarities and differences between the images from a photograph and video recording. In osteological studies, morphological assessment of the skull and pelvis are the best known. Many forensic anthropologists rely on variation in, for example, the shape of the mastoid process or subpubic angle, as much as, if not more than, the metric features of these structures. As a matter of fact, anthroposcopic characteristics may go beyond population boundaries; for example, the subpubic angle is a better structure for determining sex than measurement of the entire pelvis.

There are two considerations in morphological analysis. A structure can be viewed as large or small with a scale in between. In some ways, this is similar to anthropometry because it can be measured with a numeric scale. Therefore, using the above example, the subpubic angle can be very wide or U-shaped or very narrow and V-shaped. The same applies to the size of the mastoid process. In this case anthropometry and anthroposcopy assess the same anatomic features. A morphological feature can also be nonmetric or discrete (not measurable). There are hundreds of such features in the human skeleton. Some of these include intersutural or wormian bones. For example, the inca bone is a structure formed where the sagittal suture meets the lambdoid suture and is not present in every individual. The third trochanter of the femur is rarely seen in a skeleton These two are examples of features that are often assessed as 'present' or 'absent' when a skeleton is examined. Because of this simplicity, some anthropologists have thought such features might aid the study of genetics in skeletal remains.

There are also morphological features that are not as simple as these nonmetric traits but assessment of which may depend on shape alone. For example, accessory articular facets on the ilium and the corresponding side on the sacrum are morphological features and provide considerable information about an individual's lifestyle, even though they are rarely measured linearly or otherwise.

Discussion

Anthropometry and anthroposcopy have been the most important research tools in biological and forensic anthropology. These two methods of observation and data collection can be made both on the living and on skeletonized human remains. As research tools, they have contributed to the analysis of human variation in terms of race, sex and body dimensions, such as stature. These areas of research have explained those dimensions and morphological features that describe sexual dimorphism, and the differences between sexes that may have been caused by social and physical environmental factors or simply by evolutionary mechanisms, such as selection. Many growth studies of the musculoskeletal system have been based on anthropometry of children. Predictions can be made from such studies; for example, to determine whether a child has the chromosomal mutation that will cause the Down syndrome, or what the predicted height will be at a given age.

Forensic anthropological anthropometric (and osteometric) dimensions have been used to develop many discriminant function formulae to determine sex and race from the skeleton. They are also used to estimate stature from long bones. These and other uses assist the police in identifying both the unknown victim and the culprit who may have committed the crime. The same dimensions in the study of a video recording of a crime scene, facial size and proportions of facial features can be estimated: for example, the length of a person's nose can be estimated from a video image and a photograph. Some anthropologists even attempt to estimate stature and the proportions of one part of the body to another in video surveillance images recorded during the commission of a crime.

Morphological analysis of unmeasurable features usually falls into the area of anthroposcopy. These features are assessed qualitatively (without using any measuring device) by an experienced person. While some anthropologists use models to compare morphological features, experience is extremely important because of the geographic diversity of the human species and resulting differences between them. Its use in forensic anthropology is mainly seen in the analysis of the skeletal remains rather than living individuals. An exception to this may be the evaluation of facial images in a photograph and video tapes. To these one may add the study of growth-related changes in the human body, especially when there is the question of age estimation from the picture of a person who was thought to be a child. Morphological assessment of osteological remains are important areas of observation in forensic osteology. It provides not only features of sex and race but also shows variation in the human skeleton that may be explained in terms of asymmetry, pathology or anomaly. Each of these features may provide information about age, pathology and trauma, time since death, effects of animals on the skeleton, and information as to whether a particular feature was caused ante mortem, perimortem or post mortem.

See also: **Identification/Individualization**: Overview and Meaning of ID. **Anthropology**: Skeletal Analysis; Morphological Age Estimation; Determination of Racial Affinity; Assessment of Occupational Stress; Skeletal Trauma, **Facial Identification**: Photo Image Identification.

Further Reading

Acsádi G and Nemeskéri J (1970) *History of Human Life Span and Mortality*. Budapest: Akadémiai Kiadó.

Bowles GT (1977) *The People of Asia*. New York: Scribner.

Bräuer G (1988) Osteometrie. In: Knußmann R (ed.) *Anthropologie: Handbuch der Vergleichenden Biologie des Menschen*. Vol. 1: *Wesen und Methoden der Anthropologie*. Part 1: *Wissenschaftstheorie, Geschichte, Morphologishe Methoden*, pp. 160–232. Stuttgart: Fischer.

Coon CS and Hunt EE (1965) *The Living Races of Man*. New York: Knopf.

El-Najjar MY and McWilliams KR (1978) *Forensic Anthropology: The Structure, Morphology, and Variation of Human Bone and Dentition*. Springfield, IL: Thomas.

Giles E and Elliot O (1962) Race identification from cranial measurements. *Journal of Forensic Sciences* 7:147–157.

Hauser VG and De Stefano GF (1989) *Epigenetic Variants of the Human Skull*. Stuttgart: Schweizerbart'sche.

Hiernaux J (1974) *The People of Africa*. London: Weidenfeld and Nicolson.

Hrdlička A (1939) *Practical Anthropometry*. Philadelphia: Wistar Institute.

İşcan MY (1983) Assessment of race from the pelvis. *American Journal of Physical Anthropology* **62**(2):205–208.

İşcan MY (1993) The introduction of techniques for photographic comparison: potential and problems. In: İşcan MY and Helmer RP (eds) *Forensic Analysis of the Skull: Craniofacial Analysis, Reconstruction, and Identification*, pp. 57–70. New York: Wiley-Liss.

İşcan MY and Ding S (1995) Sexual dimorphism in the Chinese femur. *Forensic Science International* **74**:79–87.

İşcan MY and Miller-Shaivitz P (1984) Determination of sex from the femur in blacks and whites. *Collegium Antropologicum* 8(2):169–175.

İşcan MY and Steyn M (1999) Craniometric assessment of population affinity in South African blacks and whites. *International Journal of Legal Medicine* **112**(2):91–97.

Knußmann R (1988) *Anthropologie: Handbuch der Vergleichenden Biologie des Menschen*. Vol. 1: *Wesen und Methoden der Anthropologie*. Part 1: *Wissenschaftstheorie, Geschichte, Morphologische Methoden*. Stuttgart: Fischer.

Krogman WM and İşcan MY (1986) *The Human Skeleton in Forensic Medicine*. Springfield, IL: Thomas.

Martin R and Saller K (1957) *Lehrbuch der Anthropologie*, vol. 3. Stuttgart: Fischer.

Saunders SR (1989) Nonmetric skeletal variation. In: İşcan MY and Kennedy KAR (eds) *Reconstruction of Life from the Skeleton*, pp. 95–108. New York: Wiley-Liss.

Sjøvold T (1973) The occurrence of minor non-metrical variants in the skeleton and their quantitative treatment for population comparisons. *Homo* **24**:204–233.

Trotter M (1970) Estimation of stature from intact limb bones. In: Stewart TD (ed.) *Personal Identification in Mass Disasters*, pp. 71–84. Washington, DC: National Museum of Natural History.

Archaeology *see* **Anthropology**: Archaeology; Excavation and Retrieval of Forensic Remains.

Arson *see* **Fire Investigation**: Chemistry of Fire; Evidence Recovery at the Fire-scene; Fire-scene; Fire-scene Patterns; Laboratory; Physics/Thermodynamics; Types of Fire.

Asphyctic Deaths *see* **Causes of Death**: Asphyctic Deaths.

Asphyxiation *see* **Autoerotic Death**.

AUTOEROTIC DEATH

B E Turvey, Knowledge Solutions LLC, Watsonville, CA, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0740

Introduction

Sexual fatalities are deaths that occur as the result of, or in association with, sexual activity. They include deaths that occur during consensual sexual behavior (accidental, natural and homicidal), deaths that occur during criminal sexual homicides (homicidal), and deaths that occur as a result of autoerotic behavior (most commonly accidental). In order to serve justice, it is important that medicolegal death investigators have the tools to discriminate between such cases. This can be accomplished by understanding the nature of autoerotic behavior, investigating death scene characteristics, and the performance of psychological autopsy techniques.

Autoerotic behavior is any act that is sexually self-gratifying. An autoerotic death occurs during autoerotic behavior in which a device, apparatus, prop, chemical or behavior that is engaged to enhance sexual stimulation causes death. Conservative estimates suggest that between 500 and 1000 autoerotic fatalities occur each year in the United States alone. Such deaths are most commonly accidental in nature. There are certainly some cases in which the victim clearly intended to commit suicide by autoerotic means, but these cases are very rare. When an autoerotic fatality has been identified, it should be assumed accidental in manner unless there is strong evidence to the contrary. Furthermore, autoerotic fatalities are not the result of criminal behavior, and should not be treated as crimes. Once a determination of autoerotic fatality has been made, law enforcement may no longer be required to investigate the matter further.

The sexual pleasure sought by individuals who engage in autoerotic activity is primarily achieved through physical stimulation and psychological stimulation. These two sources are not exclusive of each other. For example, the behavior of masturbation is physically gratifying; if the individual places a mirror so that they may observe themself masturbating, then this becomes psychologically gratifying as it fulfills a visual fantasy need.

The danger in autoerotic activity arises when an individual acclimates to their current sexual stimulus and starts to do things that put their life at risk. To achieve an increasingly enhanced state of physical or psychological pleasure, an individual may begin to employ a high-risk device, apparatus, prop, chemical or behavior that puts their life in danger. The more often an individual uses dangerous, high-risk elements during autoerotic behavior, the greater the

chance that their practices will result in an autoerotic death.

Table 1 lists examples of high-risk devices, props and chemicals found in published cases of autoerotic death. The term high-risk may be subjective to the health of the victim. The term device or apparatus refers to any item that an individual uses or adapts for physical pleasure. The term prop refers to any item that an individual uses or adapts for psychological pleasure (fantasy). Items can be both a device/apparatus and a prop (devices can easily become fetishized over time), and items can be used in combination with each other. The term chemical refers to any chemicals or drugs used that are intended to enhance sexual arousal during autoerotic behavior.

Autoerotic Asphyxia

The most common form of autoerotic death results from autoerotic asphyxia. Asphyxial deaths are caused by a restriction of oxygen to the cells. There are three general categories of asphyxial deaths: suffocation, strangulation and chemical asphyxia. Autoerotic asphyxia most commonly takes the form of self-induced hanging, including deaths caused by or in association with strangulation and chemical asphyxia. Hanging is a form of strangulation where death is caused by compression of the blood vessels in the neck. The compression results in an insufficient amount of oxygenated blood reaching the brain.

This pattern of repetitive sexual self-gratification is referred to as asphyxiophilia, hypoxophilia or sexual asphyxia. It is also referred to as autoerotic asphyxia syndrome. Those who engage in this dangerous, repetitive pattern of sexual behavior typically intend to achieve physical pleasure. The physical pleasure comes from a temporary state of anoxia (a partial reduction of oxygen to the brain) that results from asphyxial behavior. By reducing the amount of oxygen to their brain, they induce a semihallucinogenic and euphoric state.

Sexual asphyxial behavior is often combined with fantasy-oriented behavior intended to satisfy a need for psychological pleasure. Fantasy-oriented behavior, often found in the form of props, is evidence of development and refinement of autoerotic fantasy over a period of time (suggesting that the incident is not at all isolated). Fantasy-oriented behavior can include, but is by no means limited to, the following:

- The use of pornographic material during asphyxiation;
- The repeated use of special fantasy items and objects;
- The repeated use of a special location for the asphyxial ritual;
- The use of bondage-related devices and/or complex ligature arrangements;
- In the case of males, the use of female undergarments, including cross-dressing behavior (cross-dressing during autoerotic behavior may be suggestive of transvestitism, but it is most often found to be employed for its masochistic value during the event).

Atypical Autoerotic Deaths

Nonasphyxial autoerotic deaths are often referred to as atypical autoerotic deaths because they involve sexual self-stimulation by means other than asphyxia. Recall that sexual asphyxia is the leading form of autoerotic fatality.

An atypical autoerotic death can include all manners and forms of dangerous behaviors that are intended to increase sexual gratification, either physically or psychologically. Such behavior includes things such as: self-torture/mutilation; insertion of foreign objects in the mouth, penis, vagina and rectum; extreme, restrictive bondage; the use of electricity; use of fire (self-immolation); the use of water (self-immersion); and the use of dangerous toxins, drugs and chemicals.

One example could be an individual who used a

Table 1 Examples of high-risk devices found in cases of autoerotic death

Device/apparatus	*Props*	*Chemicals*
• Asphyxial ligature (rope, belt, cord, etc.) • Complex ligature • Duct tape for mummification • Electrical wires attached to the body • Fire (sexual immolation) • Immersion in a bathtub • Plastic bag over the head • Power hydraulics • Restrictive bondage	• Firearm • Knife • Sharp, oversized or unclean fetishized objects inserted into orifices (bolts, large cucumbers, oversized dildos) • Vacuum cleaner attached to or inserted into the body	• Amyl nitrate • Cocaine • Freon • Gamma-hydroxybutyrate (GHB) • Methamphetamine • Methylenedioxyamphetamine (MDMA) (Ecstasy) • Nitrous oxide • Propane • Tetrachloroethylene

Figure 1 Photograph showing multiple grooves worn into the beam with rope in situ. Demonstrates evidence of prior autoerotic activity of a similar nature.

great deal of cocaine while watching a pornographic video who, while masturbating, died of a heart attack. Another example could include an individual who repeatedly inserted a large, pointed carrot into their rectum, causing a bowel perforation that led to an infection, and then death. Still another example could include an individual who attached electrodes to their anus while masturbating and died as the result of cardiac fibrillation. A final example could include an individual who repeatedly immersed themselves in bathtub water, and then accidentally bumped their head and drowned. In that case, a determination of autoerotic aqua-eroticism would be possible if there were evidence of masturbation at the time of death, and a diary written by the victim detailing the heights of sexual pleasure achieved during aqua-erotic behavior in the past.

Death Scene Characteristics

As part of the medicolegal investigation into any death that occurs in association with sexual activity, it is important to establish the presence of key death scene characteristics before it can be appropriately classified as an autoerotic fatality. In doing so, death investigators can hope to differentiate autoerotic death scenes from other similar types of death scenes such as suicides and sexual homicides.

The particular death scene characteristics discussed in this section are often ignored or overlooked entirely by investigators. The two main reasons for this are the general lack of knowledge about the nature of autoerotic behavior, and the fact that autoerotic fatalities often share common death scene characteristics with both suicides and sexual homicides.

There are typically 12 general behavioral characteristics that investigators can look for to help them identify autoerotic death scenes. It is not necessary for all the behavioral characteristics listed to be present for a determination of autoerotic fatality. These 12 items are not used as an absolute, inflexible checklist. These behavioral characteristics are intended to be a guide that helps investigators identify various forms of autoerotic behavior within a crime scene. The search for these characteristics is the basis for questions that death investigators consider when determining whether or not an autoerotic fatality occurred.

- *Location*: a secluded area with a reasonable expectation of privacy, i.e. a locked bedroom, bathroom, basement, attic, garage, workshop, motel room, or wooded area etc.
- *Body position*: in asphyxial hanging deaths, the victim's body may be partially supported by the ground, or the victim may even appear to have simply been able to stand up to avoid strangulation.
- *High-risk elements*: these are items that are brought in to the autoerotic activity which enhance physical or psychological pleasure. They increase the risk of autoerotic death (see Table 1).
- *Self-rescue mechanism*: this is any provision that allows the victim to voluntarily stop the effect of the high-risk element, i.e. literature dealing with escape mechanisms, a slip knot in a ligature around the victim's neck, the freedom to punch a hole in a

Figure 2 Photograph showing suspension in the chair and bindings to the legs, and pantyhose on ground used as gag, which is evidence of bondage activity. Also shows decedent wearing pantyhose and high-heeled women's shoes, which is evidence of fetishistic cross-dressing. Note exposed genitals and vibrator used to assist sexual gratification.

plastic bag sealed over the victim's face, a remote safety button on a power hydraulic in or near the victim's reach, keys for locks, or the ability to simply stand up and avoid asphyxiation altogether.

- *Bondage*: this refers to the use of special materials or devices that physically restrain the victim. These items have a psychological/fantasy significance to the victim. The presence of this characteristic can lead investigators to believe that a death was homicidal when it was not. In cases of autoerotic death, it is important to establish that a victim could have placed the restraints on themselves, without assistance, i.e. literature dealing with bondage, items such as handcuffs, leather harnesses, wrist manacles, elaborate ligature configuration, etc.
- *Masochistic behavior*: inflicting physical or psychological (humiliation) pain on sexual areas of the body, or other body parts. It is important not only to look for indicators of current use, but of healed injuries suggesting a history of behavior when appropriate, i.e. literature dealing with sadomasochism, items such as a spreader bar between the ankles, genital restraints, ball-gag, nipple clips, cross-dressing, suspension, etc.
- *Clothing*: the victim may be dressed in fetishistic attire, or they may be dressed in one or more articles of female clothing. However, cross-dressing or fetishistic attire may be absent. Clothing is not always a useful indicator in cases of autoerotic death. It is possible for victims of autoerotic fatalities to be fully dressed, nude, or in a state of partial undress.
- *Protective measures*: the victim often will not want injuries sustained during regularly occurring autoerotic behaviors to be visible to others. Injuries may be inflicted only to areas that are covered by clothing, or they may place soft, protective material between their skin and restraining devices and/or ligatures to prevent abrasions and bruising, i.e. a towel placed around the victim's neck, beneath a hanging ligature, wrist restraints placed over the victim's clothing, etc.
- *Sexual paraphernalia and props*: items found on or near the victim that assist in sexual fantasy, i.e. vibrators, dildos, mirrors, erotica, diaries, photos, films, female undergarments, method for recording the event (positioned audio tape, videotape, camera), etc.
- *Masturbatory activity*: the absence of semen from the scene is not a conclusive indicator of autoerotic death. The victim may or may not have been manually masturbating at the time of death. Evidence of masturbation strongly supports a determination of autoerotic death, however, and may be suggested by the presence of semen, tissues, towels and lubricant on hands and sexual areas.
- *Evidence of prior autoerotic activity*: this includes evidence of behavior similar to that found in scene that pre-dates the fatality, i.e. permanently affixed protective padding, plastic bags with repaired 'escape' holes, pornography from many different dates, a large collection of erotica, complex high-risk elements (very complex ligature configurations), complex escape mechanisms, healed injuries, grooves worn in a beam from repeated ligature placement, homemade videotape of prior autoerotic activity, witness accounts of prior autoerotic behavior, etc.
- *No apparent suicidal intent*: the victim plans for future events in their life. Absence of a suicide note is not an indication of an autoerotic event. If one is present, it must be determined that it was written around the time of death, and is not a prop (part of a victim's masochistic fantasy), i.e. the victim made plans to see close friends or go on trips in the near future, no history of depression, recently paid monthly bills, spoke to friends of looking forward to a specific event, etc.

As stated, not all of the above characteristics need be present. The characteristics that must be present for a determination of autoerotic death are, at the very least:

- Reasonable expectation of privacy;
- Evidence of solo sexual activity;
- Evidence of prior high-risk autoerotic practice;
- No apparent suicidal intent.

In terms of victim sex, the autoerotic death scenes of males tend to be quite different from the autoerotic death scenes of females. In male and female cases alike the occurrence of autoerotic behavior has been found to be both secretive and repetitive.

It has been found, however, that male victims tend to use a far greater range of elaborate devices and props during autoerotic behavior. These props are often designed to cause real or simulated pain. There is also a greater occurrence of cross-dressing and use of pornographic material.

Female victims, on the other hand, tend to be found naked, often using only a single device, with no excessive or elaborate apparatus or props. It has been repeatedly suggested that these death scene differences may account for the high level of under-reporting and/or mislabeling of female autoerotic deaths. This is especially true when female victims of autoerotic fatalities are deeply involved with self-binding and other sadomasochistic behavior. This scenario can readily appear to investigators as a sexual homicide, perhaps the work of a sexual predator.

Other factors that tend to influence the characteristics found at an autoerotic death scene include the friends and family of the victim. When friends or family members discover a loved one that has been the victim of an autoerotic fatality, very often their first impulse is to clean up the scene. They remove and dispose of potentially embarrassing items (i.e. pornography, articles of female clothing, sexual devices), and generally act to preserve their loved one's dignity before notifying authorities. This is not criminal behavior done with criminal intent, and should not be treated as such. However, investigators must be aware that this type of intervention can and does happen, and they must anticipate dealing with it.

Psychological Autopsy

An important part of any medicolegal investigation into a death that occurs in association with sexual activity is the performance of what is referred to as a psychological autopsy. Psychological autopsy is a term that refers to the process that death investigators and mental health experts use, in collaboration, to determine the state of mind of a person before they died.

The question that most commonly arises in cases involving a person who was alone at the time of death is whether or not death was an accident or a suicide. It is believed that by making a very thorough examination of the victim's lifestyle and history (victimology), a more accurate determination can be made as to whether a death was an accident or a suicide. The question is not always answerable, but it most certainly cannot be answered without detailed information about both the death scene and the victim's history.

There are three major areas of investigative consideration when performing a psychological autopsy: (1) wound pattern analysis; (2) victim state of mind; and (3) victim mental health history.

To evaluate wound patterns on the victim, a complete analysis is required. The injuries should be examined in the crime scene (in the environment in which they were created) and at autopsy (after they have been cleaned up) when possible. There should also be, at the very least, an examination of all documentation relating to the crime scene and the autopsy (sketches, photos, videos, reports, notes, etc.). Key questions that are asked of the wounds include whether or not there are hesitation marks, whether or not the deceased could have caused the injuries, where any weapons that were used were found, and the precise causes of death.

To evaluate the victim's state of mind prior to their death, a complete background study of the victim's history and personality (victimology) is required. This includes conducting extensive interviews of the victim's friends and family, including close personal friends and any current or past intimates. Death investigators look for any of the classic warning signs of suicidal behavior. These warnings signs include giving away possessions, sudden cheerfulness or emotional calm after a prolonged depression, speaking of life in the past tense, social isolation, extreme changes in regular habits or activities, recent stressful events, an accumulation of stressful events, or recent deaths of friends and/or family. If there is a suicide note, this should be analyzed and compared to a sample of the victim's handwriting by a questioned document examiner. If a note is not found, handwriting samples should be taken anyway in case a note is found later. If a note is determined to have originated from the victim, then the contents of the note may reveal their state of mind and possible motives for their behavior.

To evaluate a victim's mental health history, a full and complete analysis has to be made of their medical and mental health interactions. This means taking into account whether or not the victim was or had been under any professional care, whether or not they were on any medication, and whether or not they had attempted suicide in the past.

These types of intensive background studies of the victim by law enforcement or medicolegal death investigators are not routinely made in sexual fatalities of any type (including sexual homicides), and as a result many autoerotic deaths are unreported or mislabeled as homicidal, suicidal or undetermined deaths. When confronted with a potential sexual fatality, it is requisite that investigators understand and appreciate the characteristics of autoerotic fatalities, and that they employ the use of psychological autopsies to understand and evaluate the behavioral evidence in the case.

See also: **Causes of Death**: Asphyctic Deaths. **Crime-scene Investigation and Examination**: Scene Analysis and Reconstruction.

Further Reading

Adelson L (1973) *The Pathology of Homicide*. Springfield: Charles C Thomas.

Burgess A, Dietz P and Hazelwood R (1983) *Autoerotic Fatalities*. Lexington, Mass: D.C. Heath.

Demartino M (1979) *Human Autoerotic Practices*. New York: Human Sciences Press.

DiMaio D and DiMaio V (1993) *Forensic Pathology*. Boca Raton: CRC Press.

Geberth V (1996) *Practical Homicide Investigation*, 3rd edn. New York: CRC Press.

Hingsburger D, Money J and Wainwright G (1991) *The Breathless Orgasm: A Lovemap Biography of Asphyxiophilia*. New York: Prometheus Books.

Holmes R (1991) *Sex Crimes*. London: Sage Publications.

Krafft-Ebing R von (1997) *Psychopathia Sexualis*. London: Velvet Publications.

Automobiles *see* **Accident Investigation**: Airbag Related Injuries and Deaths; Driver Versus Passenger in Motor Vehicle Collisions; Motor Vehicle.

B

Ballistics *see* **Firearms**: Humane Killing Tools; Laboratory Analysis; Range; Residues; Weapons, Ammunitions and Penetration.

BASIC PRINCIPLES OF FORENSIC SCIENCE

B D Gaudette, Royal Canadian Mounted Police, Ottawa, Canada

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0419

Introduction

> The view is widely held that there is no unity to forensic science. Eminent forensic scientists have themselves complained that forensic science is seen merely as an application of the knowledge generated by other sciences, lacking any underlying theory or principles. This is illustrated by the fact that many working forensic scientists regard themselves as chemists or biochemists. Apart from anything else this creates the difficulty that when the prosecution calls a forensic scientist the defense will frequently be able to retain a more highly qualified chemist, biochemist, or geneticist.
>
> We argue that forensic science does indeed have a unity. This is because it is concerned with drawing inferences relating to an event that is the concern of a court from individual items of non-replicable data. Scientists and statisticians working in other fields use techniques that do not provide and, in many cases, deny the possibility of answers to such questions. Forensic scientists therefore should think in such terms, should qualify themselves in court as forensic scientists and should more actively question the qualifications of other scientists to give evidence. (Robertson and Vigneaux, 1995.)

This quotation sums up a problem that has plagued forensic science since its inception. As much of forensic science arose from other branches of science, principles from these other fields were often appropriated without sufficient consideration of the special constraints of forensic science. Indeed, most forensic science textbooks jump into a discussion of analytical techniques without any mention of basic principles.

Forensic science is concerned with the reconstruction of unique events – attempting to determine what happened, how it happened, where and when it happened, and who was involved. Each crime occurs under a different combination of circumstances affected by an enormous number of variables, and is thus nonreplicable. Furthermore, forensic scientists must deal with samples that are severely limited in size and quality, and which have an unknown (and often unknowable) history. In addition, the legal process imposes unique constraints and characteristics on forensic science. To accommodate all the preceding considerations, forensic science requires its own set of principles.

The Five Stages of the Physical Evidence Process

The work of forensic scientists can be best appreciated by considering the stages of the physical evidence process shown in **Fig. 1**. Forensic scientists are not present for the first stage, and an individual

Figure 1 The stages of the physical evidence process.

forensic scientist is generally not personally involved in all of the remaining stages. Nevertheless, it is critical that the forensic scientist have an understanding of, and give consideration to, each stage.

The foundation is laid when evidence is produced: when a signature is forged; when glass particles are transferred from a broken window to a pair of gloves worn by a burglar; when a victim's blood is shed on a knife used as a murder weapon and persists through an imperfect attempt at washing it off. Knowledge of how evidence occurs is crucial to understanding how it may best be recovered and interpreted.

Recovery of evidence builds on this foundation. Sometimes evidence recovery is simple and straightforward, such as seizing pieces of broken plastic from an automobile at a hit and run accident; at other times it is part of a routine practice, such as removing the liver of a poisoning victim during autopsy; at still other times, special techniques are required, such as taping to recover fibers from a break and enter point, or special lighting to render footprints or stains visible. At a crime scene, the location and relative position of evidence can play a crucial role in reconstructing an event. Special steps must be taken to record and document such information.

Recovered evidence must then be analyzed. It is this part of forensic science that is most like conventional disciplines such as analytical chemistry. However, even here unique problems can arise because the samples a forensic scientist must deal with have an unknown history and are often degraded, contaminated or otherwise environmentally challenged. Additionally, the amounts present are often several orders of magnitude less than the ideal amounts for which standard clinical or analytical methods were developed. Accordingly, forensic scientists must modify standard analytical protocols to fit their special needs and must be ready to develop their own methods when dictated by unique case circumstances.

Interpretation, the fourth phase of the physical evidence process, is the heart of forensic science. On the basis of the *results* of an examination, a forensic scientist draws a *conclusion* which he or she then interprets in giving an expert *opinion*. Such opinions can be as to the effect of a certain combination of prescription drugs and alcohol on a person's ability to operate a motor vehicle; or as to the likelihood that a particular firearm fired a bullet found at a crime scene; or as to the likelihood that a child was the biological offspring of a certain man and the daughter of his ex-best friend. In order to provide a proper interpretation of evidence, a forensic scientist needs a full understanding of a case and the circumstances surrounding it. In the early days of forensic science, it was felt that the forensic scientist should work in isolation from the rest of the investigation, as knowledge of case details would remove a forensic scientist's objectivity. Today it is recognized that interpretations can only be rendered in context, and that true objectivity is not possible in forensic science (see below).

The preceding parts of the physical evidence process are summarized in the presentation stage. Presentation most often takes the form of a laboratory report and may also involve court testimony as an expert witness. Since laboratory reports are the basis for many important decisions (whether to lay a charge, what charge to lay, who should be charged, whether to plead guilty or not guilty, whether to plea bargain, etc.) made before a case ever goes to trial, they should include all information likely to be required by police, prosecutors and others involved in the decision making process. All assumptions should therefore be clearly spelled out.

The basic principles of forensic science are linked to the stages of the physical evidence process and will be discussed in that framework. Before we begin, it is important first to consider the topic of subjectivity.

Forensic Science is of Necessity Somewhat Subjective

Is forensic science an objective science? The dictionary defines 'objective' as 'free from or independent of personal feelings, opinion, prejudice, etc.'. We have just seen that forming an opinion is one of the components of interpretation. Full knowledge of case circumstances is a key to proper interpretation, which is at the heart of forensic science. We have also seen that all forensic samples are not predictable and amenable to standardized analytical protocols. Finally, we have seen that forensic science is concerned with reconstruction of unique events. The use of statistics derived from seemingly objective databases to illuminate such unique events is dependent on numerous assumptions, all of which depend on the subjective judgment of the forensic scientist. As a result of the preceding factors, forensic science is of necessity somewhat subjective. Subjectivity has developed a negative connotation in the modern world, but it must be remembered that the value of objectivity, like beauty, is in the mind of the beholder. Furthermore, the boundary between objectivity and subjectivity is itself subjective. To be truly objective, one must discount past experience and context. Past experience and case circumstances are factors that forensic scientists use to great advantage in forming expert opinions. To be sure, forensic scientists need to be objective in the sense of being impartial, and giving due consideration to alternative hypotheses.

However, it must be remembered that objective forensic science can only exist within a framework of subjective judgement.

Principles Concerning Occurrence of Evidence

Locard exchange principle: 'The dust and debris that cover our clothing and bodies are the mute witnesses, sure and faithful, of all our movements and all our encounters.'

Corollary: The absence of evidence is not necessarily evidence of absence.

The Locard exchange principle, first formulated by French forensic scientist Edmond Locard, is the foundation of forensic science. It is often worded as 'every contact leaves a trace', and means that whenever two objects come into contact there will always be a transfer of material from one to the other. Depending on the nature of the objects and the extent of contact, the amount of material transferred may be very small (only a few molecules), too small to be detected even by today's sophisticated methods. Also, depending on the amount of time that passes, the nature of the recipient object, and the environment and actions to which it is subjected, much or all of the transferred material may be lost prior to the recovery stage. It is these factors that lead to the corollary. Just because no transferred material can be detected does not necessarily mean that no contact occurred. As with everything else in forensic science, context and unique case circumstances are all important. If a suspect vehicle apprehended 15 minutes after a high speed hit-and-run pedestrian accident failed to show any indication of adhering blood, tissue, hairs or fibers, it could be safely concluded that it was not involved. However, if no hairs or fibers were found on an assault suspect's clothing seized 2 weeks after the offense and after he had washed it several times, no inference about lack of involvement could be made. Knowledge of the Locard exchange principle coupled with information about case circumstances can focus the search for evidence. Each new technological development which increases the sensitivity of analytical methods used by forensic scientists also increases the importance of the Locard exchange principle. A few years ago, forensic scientists would not even have attempted to look for evidence that was not readily visible. Today, however, the enhanced sensitivity of modern methods is exemplified when, through DNA analysis of a few epithelial cells transferred during the contact, forensic scientists can sometimes establish that a suspect handled a knife used as a murder weapon.

Principles Concerning Evidence Recovery

Principle of evidence recovery: First, do no harm. Nothing should be added, damaged or obliterated in the recovery process.

Since there are no formal forensic science principles concerning evidence recovery, it is helpful to appropriate a basic principle from the field of medicine. The most important principle that should be borne in mind by those attempting to collect evidence is that nothing should be added, lost, damaged or obliterated in the recovery process. Particular attention should be paid to avoiding contamination, a concern that gains increasing importance with each advance in analytical sensitivity. Where there is risk of losing or damaging evidence, great care should be taken and the appropriate experts should be called in. Exhibit items should be safely and securely packaged as soon as possible. If an object is at all portable it should be transferred to the laboratory for evidence recovery under controlled conditions. An extension of the principle of doing no harm is that those collecting exhibit items should ensure that they do no harm to themselves. Crime scenes and recovered evidence may pose biological or chemical hazards. Appropriate health and safety measures must be taken when collecting and transporting evidence.

Principles Concerning Analysis

Principle of analysis: Use the scientific method.

The analysis stage is that part of forensic science which most closely parallels other sciences. As such, the fundamental principle is the scientific method. The steps of this method (observation, data collection, conjecture, hypothesis, testing results, theory) are followed, often unconsciously, in answering scientific questions. For example, in attempting to reconstruct events which occurred at a crime scene, a forensic scientist carefully observes the scene and gathers all known facts. This then leads to conjecture, which is refined to form hypotheses, which are then tested to see how well they conform to the known facts. As additional facts are revealed through further observation or experimental testing, it often becomes possible to develop a theory of what took place.

In applying the scientific method, forensic scientists use both inductive and deductive reasoning.

Novel forensic methodologies must undergo developmental validation following the scientific method to ensure the accuracy, precision and reproducibility of the procedure. Two types of evidence may be different in nature but may share analytical methodology and thought processes.

Based on observations and data, forensic scientists must first classify or identify recovered evidence by placing it into a group of items with similar characteristics. For instance, firearms can be classified according to their caliber and rifling characteristics. Such groups can have varying degrees of specificity; for example, 'fibers', 'polyester fibers', 'pink polyester fibers', 'pink heavily delustered polyester fibers', etc. Once an item has been classified, it (or an impression made from it) can be compared to another item (or an impression made from it) from either a known source or a legal standard (such as an allowable blood alcohol level or a definition of a prohibited drug or restricted weapon) to test the hypothesis that it matches.

It is only valid to compare items from the same class. For example, in microscopic hair comparison questioned human scalp hairs must be compared to known human scalp hairs and not to human pubic hairs or wool fibers. (Note, however, that it is important to define carefully what is being compared. For example, DNA from scalp hairs can be compared to DNA from pubic hairs or from blood stains.) Comparisons can be visual (e.g. microscopic comparison of morphological hair characteristics or of toolmarks), graphical (infrared spectra of paints) or numerical (refractive indices of glass samples, melting points of synthetic fibers, etc.).

In conducting comparisons, forensic scientists must compare both 'class characteristics', those characteristics arising from manufacture or natural production, and 'accidental characteristics', those characteristics acquired through use or abuse. (With living things, the term 'individual characteristics' is generally used instead of 'accidental characteristics'. Individual characteristics arise during development and environmental exposure.) Since all members of a group will share class characteristics, having such matching characteristics is a necessary but not sufficient step towards individualization. Provided they are sufficiently unusual or numerous, matching accidental or individual characteristics can, on the other hand, lead to individualization. An item is said to be individualized when it is matched to only one source and the forensic scientist is morally certain that another matching item could not occur by chance. Individualization is based on discernment of traits that are so rare, either alone or in combination with other traits, that it is unreasonable to expect that they could be duplicated by chance alone.

Principles Concerning Interpretation

Principle of individuality: Two objects may be indistinguishable but no two objects are identical.

Principle of comparison: Two objects are said to match (be indistinguishable) when there are no *unexplained, forensically significant* differences between them.

The ultimate goal of a forensic scientist is to be able to uniquely identify or individualize an object. Items that are naturally occurring are, at least in theory, unique and potentially amenable to individualization. (As will be seen below, individualization of mass produced manmade items is much more problematic, both in theory and practice.) As noted above, however, forensic science generally involves *comparisons*. If no two objects are identical in all minute details we will never have a perfectly exact comparison. How is a forensic scientist to interpret this? It is here that experience and training come to the fore. In interpreting comparisons, the forensic scientist must develop a level of stringency that is neither so strict that all objects are said to be distinguishable or so nonrestrictive that obviously different items are deemed to match. The ideal level of stringency depends on the context in which the comparisons are being conducted. If a comparison is conducted as the final and ultimate test, the motto should be 'if in doubt, throw it out'. However, if the comparison is conducted as a screening prior to other tests, a motto of 'if in doubt, include' is more appropriate.

With some types of forensic examinations for which individualization is possible in theory, the applicable forensic science methods are sufficiently advanced that individualization can also be achieved in practice. Fingerprints and forensic DNA analysis are prominent examples. In some other fields, available methodology has not yet reached the potential for individualization. For example, the only available methods for comparing hairs without roots (microscopic hair comparison and mitochondrial DNA analysis) are not presently sufficiently advanced to enable individualization. If they are handmade or possess sufficient accidental characteristics, manmade items can be individualized. Examples include unusual handwriting and a piece of torn patterned fabric physically matched to a garment. However, the vast majority of manmade objects that are mass produced and lack sufficient accidental characteristics can never be individualized, regardless of how sophisticated the analytical methodology becomes. The challenge with these items, and with those naturally occurring items which cannot as yet be individualized, is to conduct sufficient tests to narrowly limit and define the class of objects that could have a similar source, and to be able to quantitatively or qualitatively express the probability of a coincidental match. Statistics and other means of assessment of evidential value become important in such 'partial individualization'.

The two key phrases in the principle of comparison are 'unexplained' and 'forensically significant'. For example, a finding that a questioned hair from a crime scene is 5 cm shorter than any hairs in a known sample from a suspect would normally be grounds to conclude that the hair did not originate from that person. However, the additional information that the suspect received a haircut just before the known sample was obtained would provide an explanation for this difference. If all other characteristics were in agreement, a forensic scientist could still offer the opinion that the questioned hair was consistent with having originated from the suspect. Determination of which differences are 'forensically significant' is based on data from experimentation (e.g. repeated forensic DNA analysis on blood samples from the same person) and experience. Instruments and forensic scientists' judgment can be 'calibrated' through proficiency testing.

Interpretation involves hypothesis testing. In arriving at an interpretation, it is often important not just to show that results are consistent with a given hypothesis but also that they are inconsistent with plausible alternative hypotheses.

Interpretation is the most common area of dispute between forensic scientists. Although they may agree on the results, two forensic scientists may provide different interpretations based on differing experience, background information, assumptions and other factors. A similar situation exists in medicine where two doctors may provide different diagnoses for the same set of symptoms.

Principles Concerning Presentation

Principle of presentation: Working within an ethical framework, a forensic scientist should fully disclose and present impartial evidence which is readily understandable and neither overstated nor understated.

It is important for forensic scientists to have and follow a code of ethics. Most forensic laboratories and professional associations (such as the American Academy of Forensic Sciences or the Forensic Science Society) have such codes, which their members must follow.

Full disclosure of all facts, assumptions, data, conclusions and interpretations should be made. The primary vehicle for this is the laboratory report. The forensic scientist should insure that his or her work notes are complete and factual, and that they are retained and available for future reference or inspection.

In presenting evidence, whether in written reports or in verbal testimony, a forensic scientist should remain impartial. He or she should not be an advocate for either side of the case at hand; it is, however, to be expected that a forensic scientist will be an advocate for his or her opinion. Even here, however, the forensic scientist should be prepared to change an opinion whenever background circumstances or assumptions change, or new information becomes available.

The best analyses and interpretations in the world are of no use if those reading or hearing the presentation cannot understand it, or, worse, misunderstand it. A forensic scientist must be able to present complex technical issues in such a way that lay people can easily understand them. At the same time, wording and presentation must be precise enough to avoid misunderstanding.

A forensic scientist who is aware of the basic principles concerning evidence, occurrence, recovery, analysis, interpretation and presentation should be able to deliver a high quality product in casework analysis. Knowledge of these basic principles is also essential to those doing research to advance forensic science. To those teaching forensic science, the basic principles represent a foundation upon which more detailed and practical knowledge can be laid.

See also: **Ethics**. **Evidence**: Statistical Interpretation of Evidence/Bayesian Analysis. **History**: Forensic Sciences.

Further Reading

DeForest P (1999) Recapturing the essence of criminalistics. *Science and Justice* **39**:196–208.

DeForest P, Gaensslen R and Lee H (1983) *Forensic Science: An Introduction to Criminalistics*. New York: McGraw-Hill.

Evett I and Weir B (1998) *Interpreting DNA Evidence*. Sunderland, MA: Sinauer.

Gaudette BD (1986) Evaluation of associative physical evidence. *Journal of the Forensic Science Society* **26**:163–167.

Inman K and Rudin N (1997) *An Introduction to Forensic DNA Analysis*. Boca Raton, FL: CRC Press.

Locard E (1930) The analysis of dust traces. *American Journal of Police Science* **1**:276–291.

Robertson B and Vignaux GA (1995) *Interpreting Evidence*. Chichester: Wiley.

Bayesian Analysis *see* **Evidence**: Statistical Interpretation of Evidence/Bayesian Analysis.

C

Capillary Electrophoresis *see* **Analytical Techniques**: Capillary Electrophoresis in Forensic Science; Capillary Electrophoresis in Forensic Biology.

Cars *see* **Accident Investigation**: Motor Vehicle; Airbag Related Injuries and Deaths; Driver Versus Passenger in Motor Vehicle Collisions.

CAUSES OF DEATH

Contents

Overview
Asphyctic Deaths
Blunt Injury
Burns and Scalds
Postmortem Changes
Scene of Death
Sharp Injury
Sudden Natural Death
Systemic Response to Trauma
Traffic Deaths

Overview

A Tracqui, Institute de Médicine légale, Strasbourg, France

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0423

Introduction

The purpose of death investigation is usually to answer the following questions: Who died, where, when, why and how? Thus the question of the cause, manner and mechanisms of death (the why and the how) is of central importance for every forensic practitioner. This introductory article aims to briefly present, mainly from an epidemiological point of view, the major causes of death encountered in developed countries. Deaths of particular medico-legal relevance, e.g. homicides and suicides are more widely developed in other articles.

According to their etiology and judicial relevance, deaths may be classified as follows:

- **Natural deaths** comprise all deaths presumed or certified of medical original, occurring under non-suspicious circumstances. They usually concern people displaying a history of acute or chronic, life-threatening disease and dying in hospitals, other health facilities or at home, under the care of practitioners able to assess the cause of death.
- **Violent deaths** are all deaths not caused by the

spontaneous, terminal evolution of a pathological process, but resulting from an external (and usually sudden) action or force. They include homicides, suicides and accidents.

- **Suspicious or unattended deaths** are deaths initially of dubious origin that require further forensic investigations to be ruled as natural or violent (e.g. young adults found dead at home, or sudden death at a workplace), or deaths of obvious natural or violent origin but which occurred under suspicious circumstances (e.g. ischemic heart disease in a context of medical mishap, or firearm death without prior indication on the suicidal or criminal origin).

Overall most deaths occur from natural causes (**Table 1**). The three main causes in all developed countries are diseases of the heart, malignant neoplasms and cerebrovascular diseases. There are, however, outstanding differences in the distribution of the leading causes of death according to the age group.

- **Infant mortality** (i.e. all deaths of subjects <1 year old) is mainly linked to factors originating in pregnancy or the perinatal period: congenital anomalies, disorders related to short gestation and/or low birth weight, sudden infant death syndrome, respiratory distress syndrome, infections specific to the perinatal period.
- **Children, teens and young adults** are disproportionately affected by deaths of violent origin. As shown in Table 1, the three leading causes of death in young Americans aged 15–24 years are accidents, homicides and suicide.
- **Adults over 45 years and the elderly** are more concerned by deaths of natural origin. Over half of all deaths recorded in this age group are due to heart failure or malignant neoplasms.

Natural deaths are frequently considered of limited forensic interest. This is, however, not always the case, as some of them may raise social or legal problems (deaths of unknown individuals, deaths at work, deaths in legal custody, deaths involving medical liability, cause of death contested by insurers, sudden infant deaths) that can be solved only by forensic examination. In addition the definition of what is relevant to forensic investigation differs strongly depending on the local judicial system. In many European countries (e.g. France, Italy, Spain) postmortem examination is ordered only for certain limited categories of death and in the context of a judicial inquiry, primarily to establish whether or not there has been any criminality or possible negligence leading to the death; once these possibilities have been ruled out, it does not remain mandatory to establish the precise cause of death. In other countries

Table 1 Leading causes of death in the USA, 1996 (death rates per 100 000 population in specified group)

	All ages	*1–4 years*	*5–14 years*	*15–24 years*	*24–44 years*	*45–64 years*	*≥65 years*
Diseases of the heart	276.6	1.3	0.9	2.5	19.4	190.8	1810.0
Malignant neoplasms[a]	205.2	2.8	2.7	4.5	26.4	247.2	1140.2
Cerebrovascular diseases	60.5	0.4[b]	n.a.	0.5	4.1	28.9	416.2
Chronic obstructive pulmonary diseases and allied conditions	40.0	n.a.	0.4	0.6	n.a	23.9	270.6
Accidents and adverse effects	35.4	13.9	9.2	38.3	31.7	30.4	90.3
Motor vehicle	16.4	5.4	5.2	29.3	17.3	14.3	22.3
All other	19.0	8.5	4.0	9.0	14.4	16.1	68.0
Pneumonia and influenza	31.1	1.1	0.4	0.5	2.4	10.5	218.4
Diabetes mellitus	23.2	n.a.	n.a.	n.a.	3.0	23.6	136.4
HIV infection	14.8[c]	1.0	0.5	1.2	27.2	15.7	n.a.
Suicide	11.6	n.a.	0.8	12.1	15.0	14.4	17.3
Chronic liver disease and cirrhosis	9.5	n.a.	n.a.	n.a.	5.1	19.9	n.a.
Nephritis, nephrotic syndrome and nephrosis	9.2	n.a.	n.a.	n.a.	n.a.	n.a.	61.9
Septicemia	8.1	0.5	n.a.	n.a.	n.a.	n.a.	51.2
Homicide and legal intervention	7.8	2.5	1.3	18.1	11.1	n.a.	n.a.
Total death rates	875.4	40.6[b]	22.5[b]	95.3[b]	192.0[b]	725.0[b]	5052.8[b]

[a] Including neoplasms of lymphatic and hematopoietic tissues.
[b] 1995 statistics (1996 not available).
[c] Estimated.
n.a.: not available.

(e.g. the medical examiner system in the USA) postmortem investigations are carried out on a much broader range of death situations and the forensic specialists are compelled to indicate the precise cause and manner of death even if obviously natural and without any legal relevance. In addition to individual case investigation and management, this systematic approach has an epidemiological interest since it allows the collection of accurate data of prime usefulness for the establishment of population statistics.

Accidents

An accident is an unintentional event, due to the sudden and unexpected action of an external cause, resulting in physical damage to persons. In the USA accidents are the fifth leading cause of death (all ages) and the first cause for people aged 1–24 years. In 1995 they were responsible for 35.6%, 41.2% and 40.4% of all deaths recorded in persons aged 1–4, 5–14 and 15–24, respectively (death rates 14.5, 9.3 and 38.5 per 100 000 population, respectively); the overall accidental death rate was 35.5.

Accidents should be subdivided into two main categories: traffic accidents (or motor vehicle accidents) and nontraffic accidents. Traffic accidents are the main cause of accidental death in all developed countries. In 1993 90 523 Americans were killed in accidents and of these, 40 899 died in road accidents (45.2%). The prevalence of accidental deaths due to nonroad transport (water, air, railway) is very low compared with that of traffic accidents. Falls were the second cause of accidental death (13 141 deaths, 14.5%); death rates by age group show an increase of fall-related fatalities with advancing age, many of these death occurring in elderly persons who die of the complications following a fall. Other kinds of nontraffic accidents include (percentages referring to the 1993 US data):

- Accidental poisonings by drugs and medicines (including overdoses by drugs of abuse): 8.2%
- Accidents caused by fires and flames: 4.3%
- Accidental drowning: 4.2%
- Accidental inhalation and ingestion of objects: 3.5%
- Complications due to medical procedures: 2.9%
- Accidents involving firearms (handguns and other): 1.7%
- Accidental poisonings due to gases and vapors: 0.7%

According to the 1995 report of the North Carolina Office of the Chief Medical Examiner, of 2682 deaths ruled accidental in manner, 1557 (58.1%) involved motor vehicles. The nonmotor vehicle accidents distributed as follows: falls 30%, poisonings (alcohol + other) 16%, fires 13%, drowning 13%, exposure 2%, all other 26%.

Traffic accident death rates are generally lower in European countries, compared with the United States. In France 124 387 road accidents with physical damage were reported in 1998, responsible for 8437 deaths and 168 535 injured (mortality and morbidity rates 14.5 and 289.3 per 100 000 population, respectively).

Suicides

Suicide statistics remain approximate, but in most countries they depict a major problem of public health. Depending on sources (WHO, national data networks, associations), the estimated number of people dying by suicide in the world is between 440 000 and 800 000 per annum. According to the National Center for Health Statistics (NCHS), there were 30 903 deaths by suicide in the USA in 1996, representing 1.3% of total deaths (suicide rate 11.6 per 100 000). Overall, suicide is the ninth leading cause of death for Americans. There are about 775 000 annual attempts at suicide in the USA, i.e. roughly 25 attempts for every completion. These data are the average among industrialized nations. Among developed countries, Hungary has the highest suicide rates. France ranks fourth (about 12 500 deaths recorded in 1997); there are at least 90 000–115 000 attempts per year, i.e. a rate of 200–250 per 100 000 French residents aged over 15. In many European countries, there are more deaths by suicide than by road accidents (also due to the lower prevalence of traffic deaths). Suicide rates are generally much lower in developing countries.

In all developed countries, there are marked discrepancies between population groups, as follows.

- *Gender:* more women than men have a history of attempted suicide (at a ratio of 2:1 to 5:1), whereas more men than women die by suicide (at a ratio of 3:1 to 5:1) as they frequently choose more radical methods. In the USA, male and female suicide rates (1996) are 19.3 and 4.4 per 100 000, respectively.
- Age: suicide rates generally increase with age. US residents over age 65 have a suicide rate of 17.3 per 100 000 (1996) and those over age 85 of 20.2 (highest rate of any age group), whereas young people aged 15–24 years have a rate of 12.1. In France, suicide rates remain relatively level for people 25–65 (around 40 per 100 000), then increase sharply for the elderly (91 per 100 000 for the 75–84). However, since much fewer young

persons die of health-related causes, suicide represents everywhere a leading cause of death among adolescents (third in the U.S.A.).
- Ethnic origin: Whites display higher suicide rates than all other ethnic groups.
- Place of residence: in America, the state with the highest suicide rate in 1996 was Nevada (20.9 per 100 000), and the lowest was the District of Columbia (6.4 per 100 000).

Typical risk factors of suicide attempts include mental disorders, substance abuse, history of family violence, adverse life events such as loss of job or incarceration (suicide rates in jail exceed those of the general population by 10–30 times), exposure to the suicidal behavior of others, prior suicide attempts, firearms or pharmaceuticals available at home.

Suicide methods

Among industrialized nations most suicide attempts (> 70–80%) involve drug overdoses, mainly with psychotropics. However, the great majority of such attempts remain unsuccessful, because today's pharmaceuticals have become much safer: fatalities involving benzodiazepines, or selective serotonin reuptake inhibitors are exceptional, which previously was not the case with barbiturates or first-generation tricyclic antidepressants.

The three main methods observed in completed suicides are firearms, hanging and drug overdose (including pharmaceuticals and drugs of abuse). Guns are by far the most common method used by both men and women in the USA, accounting for 60.3% of all suicides in 1994, and 58.8% in 1996. This method is less prominent in Europe due to more restrictive regulations on firearms. Depending on the country, either hanging or firearms rank first in males, close to each other, followed by poisoning; females preferentially choose poisoning, jumping from heights or drowning. Hanging is the technique of choice for captive persons (patients of psychiatric wards, prison inmates) who do not have access to more sophisticated methods. In France, 98% of suicide deaths in prison are by hanging. In developing countries the most popular methods are hanging and poisoning (more harmful than in Europe or North America because it frequently involves highly toxic industrial or agricultural substances).

A second group of methods are less usual but not exceptional, as each accounts for 5–15% of suicides in developed countries: gas poisonings (mainly carbon monoxide), drowning, jumping from heights. All other methods may be held exceptional, none of them exceeding 2% of suicide deaths; they include:

- burning
- electrocution
- suicide by sharp force: wrist cutting (frequent in suicide attempts especially in females or adolescents, but almost never successful), arteriotomy, stabbing at throat or trunk, self-mutilation
- use of motor vehicles
- asphyxic methods other than hanging: strangulation, suffocation

Homicides

Homicide may be defined as the willful killing of one human being by another. Some discrepancies, however, exist between statistics originating from different sources, depending on whether they include (or not) in the definition: (1) deaths caused by reprehensible negligence (nonpardonable negligent manslaughter); (2) justifiable homicides (the killing of a felon by a police officer in the line of duty, or by a private citizen, e.g. in self-defense situations).

A total of 18 211 US residents were murdered in 1997, a rate of 6.8 per 100 000 population. Almost level since 1950 (4.6 per 100 000), the homicide rate doubled from the mid 1960s (4.9 in 1964) and peaked at 10.2 in 1980. It subsequently decreased (7.9 in 1985), then rose again to a peak of 9.8 in 1991. Since then the homicide rate has declined continuously to levels last seen in the late 1960s. In spite of this relative improvement, the homicide rates observed in the United States remain at the highest level among industrialized countries, due to legal (regulations on firearms) and sociological reasons. For instance there were 'only' 581 homicides in Canada in 1997, a homicide rate of 1.92 per 100 000, i.e. less than one-third of that of the United States but still higher than in most European countries (France 1.66, England and Wales 1.00). The recent decrease of homicide rates appears to be general. The 1997 Canadian data represent a 9% decrease from 1996 and the lowest rate since 1969. In France 963 persons were murdered in 1997, a 17.8% decrease from the previous year.

Characteristics of victims and offenders

Homicide statistics recorded in various industrialized nations highlight some characteristics of victims and offenders that may be considered as roughly constant.

Age Homicide primarily concerns young adults. In the United States the involvement of teens and young adults as both victims and offenders increased dramatically in the late 1980s whereas rates for older age groups declined. Since 1986 the highest homicide victimization rates concern people aged 18–24 years

(+16% annually during 1986–1991), whereas they were previously observed among the 25–34 years olds; homicide offending rates followed a similar pattern. This is confirmed by a continuous decrease of the average age of both victims and perpetrators: the average age of victims fell from 35.2 years in 1976 to 31.7 years in 1997, and that of offenders from 30.8 to 27.7 years. Similarly, Canadian residents aged 18–29 years had the highest homicide rate as victims in 1997 (3.39 per 100 000), whereas the elderly aged 60 years and older had a rate of 1.46; the lowest rate, 1.16, was for children and young people aged 18 and under.

Gender Men are most often the victims and the perpetrators in homicides. Homicide is the tenth leading cause of death of men in the United States. In 1997, male US residents were 3.6 times more likely than women to be murdered, and nine times more likely to be the perpetrator of a murder. The combination 'male offender/male victim' represented 67.9% of all homicides reported (male offender/female victim: 22.0%; female offender/male victim: 7.8%; female offender/female victim: 2.3%). These data are in accordance with a compilation of US homicide statistics from 1976 to 1994, which showed that 76% of the murder victims were male, as were 86% of the known murderers. However, when a sexual assault was associated with the homicide, 95% of the offenders were male and 82% of the victims were female.

Ethnic origin There are marked differences between American Blacks and Whites for both homicide victimization rates and homicide offending rates. Blacks are seven times more likely than Whites to be murdered, and eight times more likely to commit homicide. Most murders are intraracial; the compilation of data from 1976 to 1997 indicates that 85% of white victims were killed by Whites, and 94% of black victims by Blacks. It is difficult to compare these results with those of other countries, as many national regulations (especially in Europe) prohibit the computation of racial origin in population studies.

Relationships between victim and offender Most homicide victims know their killers. In Canada, only 13% of all victims were killed by a stranger in 1997. Two subcategories are of particular interest in the context of intrafamilial violence: spousal homicide, and infant and child homicide.

Spousal homicide Women are far more likely than men to be killed by their spouse. Over the past decade, three times more women than men have been killed by their spouse in Canada; in 1996, 78% of all spousal homicide victims were female and 22% were male. Women are also far more likely to be killed by their spouse than by a stranger; in 1997, 61 women were murdered by a current or exspouse in Canada, whereas 12 were killed by a stranger; during the period 1974–1992, a married woman was nine times more likely to be murdered by her husband than by a stranger. Some situations are frequent in families where the man shot and killed his wife, thus may be held as risk factors for such events: violent disputes, alcohol abuse, criminal record of the husband, financial difficulties, illegal possession of firearm, separation or divorce in the process or completed, large age difference between spouses.

Infant and child homicide In 1997 a total of 623 children under the age of 5 years were murdered in the United States. The race-adjusted homicide rates were 2.3 and 8.7 per 100 000 for white and black children, respectively. Most of the children killed were male. The majority of these homicides were committed by a parent: 27% by mothers, 27% by fathers, 6% by other relatives, 24% by acquaintances and only 3% by a stranger. In a similar way, 53 children under 12 (including 18 infants) were murdered in Canada in 1997, three quarters of them by a family member.

A significant proportion of homicides, especially those involving relatives, are followed by the suicide of the offenders. According to a report from the Canadian Department of Justice, 70% of men who shot and killed their wife in 1989 and 1990 subsequently turned the gun on themselves.

Homicide methods

The worldwide three most popular methods used for killing somebody are shooting, stabbing and beating. In the United States homicides are most often committed with firearms (roughly seven out of ten murders in all statistics), far below stabbing then blunt force. Of 18 211 people murdered in 1997, 9796 were killed using a handgun (53.8%), 2601 using another kind of gun (14.3%), 2355 using a knife (12.9%), 822 using blunt force (4.5%), and 2637 using other methods. Given an attack, guns are more likely to result in serious injury and death than if any other weapon is used: firearm attacks are about three times more likely to be lethal than knife attacks, and still many times more likely than attacks involving other methods. Firearm-associated family and intimate assaults are 12 times more likely to result in death than nonfirearm incidents.

In all developed countries but the United States, the

prominence of gun attacks in homicides is far less pronounced. During the past decade there has been an average of 1300 gun-related deaths per year in Canada and of these, homicides accounted for only about 15%, suicides 80% and accidents 5%. Over this period firearm homicides have remained relatively constant at about one-third of total homicides; for instance, of 633 homicides reported in 1996, 211 (33%) involved firearms. From 1985 to 1995 inclusive, the average firearm homicide rate per 100 000 population was 0.7 in Canada, compared to 5.6 in the United States.

The fourth homicide method by order of frequency is mechanical asphyxiation performed by strangulation or suffocation. Since this method generally requires a large strength difference between offender and victim, it is particularly frequent in the following offender/victim combinations: male adult/female adult, male adult/child or infant, female adult/child or infant. All other homicide methods may be considered rare to exceptional, including burning, drowning, intentional poisoning by solids, liquids or gases, electrocution, precipitation from heights and starvation.

See also: **Accident Investigation**: Motor Vehicle. **Causes of Death**: Scene of Death; Postmortem Changes; Sudden Natural Death; Blunt Injury; Sharp Injury; Burns and Scalds; Traffic Deaths. **Clinical Forensic Medicine**: Self-inflicted Injury. **Crime-scene Investigation and Examination**: Suspicious Deaths. **Pathology**: Autopsy.

Further Reading

American College of Legal Medicine (1998) *Legal Medicine*, 4th edn. St Louis: Mosby.

DiMaio VJM and Dana SE (1998) *Handbook of Forensic Pathology*. New York: RG Landes.

Gordon I, Shapiro HA and Berson SD (1988) *Forensic Medicine – A Guide to Principles*. Edinburgh: Churchill Livingstone.

Knight B (1996) *Forensic Pathology*, 2nd edn. London: Edward Arnold.

McIntosh JL (1998) *U.S.A. Suicide: 1994 Official Final Statistics*. (Prepared for the American Association of Suicidology). http://sun1.iusb.edu/~jmcintos/Suicide Stats. html.

Randall B (1997) *Death Investigation: The Basics*. Chicago: Galen Pr.

State of North Carolina – Office of the Chief Medical Examiner (1995) *1995 Annual Report*. http://thalamus.med.unc.edu/ocme/annreport/1995/.

Stone G (1999) *Suicide and Attempted Suicide: Methods and Consequences*. New York: Carroll & Graf.

US Census Bureau, The Offical Statistics (1998). *Statistical Abstracts of the United States, section 2: Vital Statistics*. http://www.census.gov/prod/3/98pubs/98statab/ sasec2.pdf.

US Department of Justice – Bureau of Justice Statistics (1998) *Homicide Trends in the United States*. http://www.ojp.usdoj.gov/bjs/pub/pdf/htius.pdf.

Wecht CH (1993) *Forensic Sciences – Law/Science – Civil/Criminal*. New York: Matthew Bender–Times Mirror Books.

Asphyctic Deaths

J A J Ferris, Department of Forensic Pathology, University of British Columbia, Vancouver, Canada

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0430

Introduction

The term asphyxia, as used in the traditional forensic pathology context, is not well defined and is often used to include conditions which are not truly asphyxial in nature. In most instances, asphyxiation refers to mechanical interference with oxygenation of the tissues but there are other conditions, including cyanide and carbon monoxide poisoning, where the interference with oxygenation occurs at a cellular level. The issues relating to the definitions of asphyxia and its effect on the human organism are comprehensively discussed in the literature. There are a number of natural diseases, including anemia and circulatory failure, for whatever cause, which can interfere with tissue oxygenation and produce cellular asphyxiation; however, it is not the purpose of this chapter to discuss these conditions. Similarly, many cases of drowning are associated with mechanical interference with tissue oxygenation but this topic will not be included in this chapter.

The physiological definition of asphyxia is complex and often means more than hypoxia. The forms of asphyxiation to be considered in this article are: exposure to gases in the atmosphere; mechanical obstruction by closure of external orifices; mechanical obstruction by interference with respiratory movements; mechanical obstruction by neck compression and sexual asphyxias.

General Autopsy Findings in Asphyxial Deaths

Because hypoxia is associated with an increase in pulmonary blood pressure and right heart failure,

many of the findings in asphyxial deaths are characteristic of right heart failure but are not of themselves diagnostic of asphyxiation. The significance of these features will vary with individual cases and may be completely absent or masked by other postmortem findings. Some of the most common findings are:

1. pulmonary edema with associated froth in the trachea and bronchi;
2. bulky, crepitant and overdistended lungs;
3. right ventricular dilatation;
4. petechial hemorrhages on the conjunctiva and facial skin, usually above the level of airway obstruction in cases of neck compression.

Petechial hemorrhages are usually pinhead size hemorrhagic spots caused by the rupture of tiny venules although sometimes described, probably incorrectly, as ruptured capillaries. They may be very few and difficult to detect but they may also be so numerous as to give the area of skin affected an apparent cyanotic appearance. They usually occur in the skin of the face, eyelids and conjunctiva in cases of mechanical asphyxiation involving obstruction to the tissues of the neck. They are also seen on the skin of the upper abdomen, chest and neck in cases of crush asphyxiation. Petechial hemorrhages beneath the visceral pleura and visceral pericardium are often referred to as 'Tardieu spots' and were regarded by Tardieu as pathognomonic of death by mechanical asphyxiation. It is now widely accepted that they can found on the surface of the viscera in a variety of deaths.

It appears that petechial hemorrhages are produced by two primary mechanisms. There is increased vascular permeability due to hypoxia as a direct consequence of the asphyxial process. Increased permeablity is also an important element of petechial hemorrhages developing after death in areas of lividity. Increased capillary and venous pressure associated with the mechanical obstruction to venous return is probably the most important mechanism determining the characteristic distribution of petechiae in cases of neck obstruction.

Petechiae are found in a variety of nonasphyxial deaths. In one series of 100 consecutive autopsies apparent petechiae were present in 33 cases, mostly unrelated to mechanical asphyxiation. The opinion has been expressed that the time has come to disregard these hemorrhages as diagnostic of mechanical asphyxia.

Petechial hemorrhages also occur as a postmortem phenomenon associated with the development of lividity. Petechial hemorrhages may be prominent within areas of lividity and often around the edges of lividity staining. These are almost certainly as a result of vascular decomposition with gravity-produced pressure inducing vascular leakage and hemorrhage.

It is not unusual during microscopic examination of tissues in cases of hypoxic deaths, to find tiny interstitial petechial hemorrhages. These are often observed within the brain in the perivascular areas and at best these microscopic features indicate terminal hypoxia. They should not be considered as evidence of mechanical obstruction to respiration.

Deaths Associated with Exposure to Gases in the Atmosphere

Although suffocation is not a specific term, it usually refers to deaths associated with a reduction of available oxygen in respired air. Oxygen may be reduced or absent from respired air or may be displaced by the presence of other gases, which are more readily absorbed by hemoglobin. Reduced atmospheric pressure, as may occur in aircraft cabin failure, or at high altitude environments may reduce the ability of oxygen to pass through the pulmonary alveolar walls. Diving accidents are often associated with exhaustion of the oxygen component of the usual nitrogen/oxygen mixture with death resulting from hypoxia. Breathing only nitrogen, which essentially functions as in inert gas, will cause rapid syncope and death.

Inert gases

Carbon dioxide, which although not of itself poisonous, may accumulate in the atmosphere at the expense of respirable oxygen and can cause hypoxia. Other gases, which may be biologically inert, may also displace atmospheric oxygen. Under such circumstances, the victims may initially suffer from headache or drowsiness and it is widely recognized that placing an individual in an oxygen-depleted atmosphere may be associated with very rapid systemic collapse and death.

Carbon monoxide

Carbon monoxide is one of the most frequently encountered toxic agents encountered in forensic practice. It has an affinity for hemoglobin that is between 200 and 300 times greater than oxygen. Therefore, very low concentrations of carbon monoxide in inspired air will rapidly displace oxygen from the red cells and lower the oxygen-carrying capacity of blood. Although there may be some direct toxic effects from the presence of carbon monoxide on tissue chemistry, the principal effect is undoubtedly the reduction in oxygen transport capacity.

Carbon monoxide combines with hemoglobin to produce the pink pigment, carboxyhemoglobin,

which can readily be identified at postmortem. The classic 'cherry pink' discoloration of the skin and tissues is usually evident when the carboxyhemoglobin saturation of the blood exceeds about 30%. Under most conditions following death, deoxygenation of hemoglobin occurs but in carbon monoxide poisoning the carboxyhemoglobin bond is stable and the body will retain its characteristic pinkish so-called flesh tone and discoloration for many hours after death even until decomposition commences.

Fatal concentrations of carboxyhemoglobin saturation will vary but a normal healthy adult will rarely die with saturations under 50% and concentrations as high as 80% are not infrequently recorded. In elderly people, the concentrations may be relatively low, often as low as 30% and in infants under one year of age, the concentrations are also similarly low and have been reported as low as 20–25% saturation.

In relatively low atmospheric concentrations of carbon monoxide, the symptoms and signs of poisoning may develop relatively slowly. Concentrations of 20–30% may be associated with nausea, headache and vomiting and as the concentration levels rise towards 40%, there may be seizure activity.

The majority of carbon monoxide poisoning cases encountered in forensic practice relate to fire deaths and accidental deaths where there has been incomplete combustion in heating systems with the slow accumulation of toxic levels of carbon monoxide in the atmosphere. Suicide cases, where vehicle exhaust fumes are used to induce fatal concentrations of carbon monoxide in inspired air, are also relatively common.

Cyanide poisoning

Historically, atmospheric cyanide poisoning has been relatively rare except in cases of judicial execution. However, cyanide is produced by the combustion of plastics and significant levels of atmospheric cyanide may accumulate in fires. Cyanide acts by linking with the ferric iron atom of cytochrome oxidase preventing the uptake of oxygen for cellular respiration and causes death by cellular asphyxiation.

Toxicological analysis of cyanide and interpretation of cyanide concentrations in blood and other tissues is fraught with difficulty. Not only are small concentrations of cyanide present in normal blood but cyanide can accumulate in stored samples of blood. However, even in ideal storage situations, cyanide may be lost from the sample within days of sample collection. Cyanide has also been found to diffuse through the tissues and from body cavities opened to cyanide perhaps as a result of trauma or fire. As a result, tissue concentrations derived from toxicological analysis of samples of blood and other tissues for cyanide have to be interpreted with caution and there is no generally accepted minimal or fatal level.

Mechanical Obstruction by Closure of External Orifices

Smothering

Smothering is usually defined as the blockage of the external air passages usually by a hand or soft fabric and a particular variety of suffocation called gagging may occur when fabric or adhesive tape obstructs the mouth and nose. In some textbooks, the term suffocation is often used to include conditions such as smothering. However, for the purposes of this article, the term suffocation is used as defined above and will be restricted to circumstances where there is a reduction to the amount of respirable oxygen. Smothering can be homicidal, suicidal or accidental. In homicidal suffocation, there is usually considerable disparity between the assailant and the victim or the victim is debilitated by disease, drugs or injury. Evidence of such disease or injury may be identified at the time of postmortem examination. Examination of the scene of death may not be helpful. Some cases of infant death previously attributed to sudden infant death syndrome (SIDS) have now been blamed on deliberate smothering by a caregiver and if the victim is incapacitated or unable to defend themselves then there may be little evidence of injury or struggle at the scene.

Smothering can occur as a result of mechanical occlusion to the mouth and nose during collapse of a work trench, in landslides such as the 1966 Welsh Aberfan disaster involving over 140 victims, in mine collapses and even in crowd crush situations.

Probably one of the commonest causes of suffocation seen in forensic practice is suicidal suffocation using a plastic bag. A plastic bag pulled over the head to cover the mouth and nose may or may not be fastened around the neck with a loosely tied ligature. This is one of the recommended methods in the 'suicide guide'. Accidental deaths by plastic bag suffocation are described in children playing with plastic containers.

The autopsy signs of smothering may be minimal. If pressure has been applied to the face around the mouth and nose there may be evidence of bruising and it may require full facial dissection to reveal such injuries. If the pressure has been maintained after death, then there may be pressure pallor within areas of lividity around the mouth and nose. When injuries are present around the mouth and nose in cases of suspected smothering, they should be view as highly suspicious and the possibility that the smothering has

been deliberate and homicidal considered. There may be no other findings. Petechial hemorrhages are more frequently absent than present. If present, petechiae may be found on the visceral surfaces of the thoracic organs or in the region of the larynx and pharynx.

Overlaying

This entity has been described as having an ancient pedigree. Reference is made in the Old Testament in the First Book of Kings when Solomon had to adjucate between two women both claiming motherhood of a child allegedly 'overlain'. In the twentieth century, cases of (SIDS) were attributed to overlaying, however this seems an unlikely mechanism of SIDS. It seems likely that overlaying is a relatively rare mechanism of death unless evidence of crush type injuries is found. In the absence of such injuries it is difficult to support a diagnosis of death from overlaying.

Postural asphyxia

Postural asphyxia is a form of smothering that may occur accidentally when an individual is incapacitated often as a result of alcohol or drug intoxication, coma from natural disease, injury or epilepsy. Such individuals are frequently overweight and the pressure of the weight of the head against a soft fabric may be sufficient to compress the nostrils and obstruct the mouth. Under such circumstances, cyanosis may be relatively prominent. However, because of postmortem de-oxygenation of blood, diagnosis of cyanosis at postmortem must be made with caution.

Choking

This is the term that usually refers to obstruction to the upper internal airways usually between the pharynx and main bronchi. A foreign body such as a large bolus of food, and small toys and other hard objects in cases of children, may completely or partially obstruct the air passages and can cause apparent very sudden death without producing any classical signs of asphyxiation. Benign and malignant tumors of the pharynx, larynx and main bronchi may also induce choking and although such lesions may be apparently relatively slow-growing obstructions, death is often relatively rapid with a very short history of respiratory difficulty. Such methods of relatively slow airway obstruction and choking are also seen occasionally with neck tumors and in cases of neck trauma where there may be large interstitial hematomas and even collapse of the laryngeal structures as a result of major trauma.

Pharyngeal and laryngeal injuries identified in such choking cases must be interpreted with caution since the obstructing object itself may cause injury. More frequently, attempts at resuscitation, intubation and removal of the foreign body often result in injuries to the pharynx piriform fossa and larynx.

In some cases of choking, death may be so rapid that the manifestations of hypoxia and asphyxia have had little time to take effect. It seems likely that these deaths may be caused by neurogenic cardiac arrest or have been accelerated by excess catecholamine release from the adrenaline response. Such cases are often attributed to vagal inhibition since it is well recognized that the larynx is a well innervated area structure. Deaths from vagal inhibition are discussed later in the section on Strangulation.

The vast majority of choking deaths are accidental. Suicides and homicides occur very rarely. The most common situation is the so-called 'café coronary'. Victims are usually intoxicated or have pre-existing neurological debilitation. Choking occurs while eating. However, the inhalation or aspiration of gastric contents of partly digested food as a primary event in an otherwise healthy adult is relatively unusual. Many of these cases may in fact be 'café coronaries' occurring as a direct consequence of a cardiac dysrythmic event. An important cause of fatal choking is the inhalation of blood associated with facial injuries, particularly fractures involving the facial bones. Accidental and homicidal injuries to the head and face associated with bleeding into the back of the pharynx and larynx may result in fatal choking as a result of the aspiration and inhalation of large amounts of blood.

In some cases of choking, resuscitation has resulted in removal of the causative obstruction and it may be require detailed dissection and examination to reveal any evidence of the primary cause for death.

Burking

This is a rare form of smothering death. The term is used to describe a procedure apparently invented by the Scottish murderers, Burke and Hare, during the 1820s. Their victim was plied with alcohol and one of the murderers knelt or sat on the chest and the other closed the mouth and nose with his hands. External injury was absent and but for the evidence of Hare, this mechanism of asphyxiation might never have been discovered.

Mechanical Obstruction by Interference with Respiratory Movements

Traumatic (crush) asphyxia

Interference with the movement of the primary muscles of respiration, namely the intercostal muscles and diaphragm, will result in fatal hypoxia. In most occasions, this is the result of accident in which the

body of the victim is subject to direct compression by a heavy weight as may occur in a rock fall in a mining accident. Traumatic or crush asphyxia is characterized by intense discoloration of the skin of the head, neck and upper chest with multiple petechial hemorrhages within this area of congestion. Such incidents are also associated with external and internal evidence of direct trauma to the chest and upper abdomen. The intense congestion of the skin and cutaneous petechiae characteristic of this mechanism of death may in part be due to obstruction to the superior vena cava, the subclavian veins and the veins of the head and neck.

In some instances, for example accidental burial in grain silos or sand/gravel pit collapses, the restriction to respiratory movement characteristic of traumatic asphyxiation may be superimposed on a smothering mechanism of death.

Traumatic asphyxia also occurs in mass disaster situations involving crushing in crowds. Such tragedies have occurred at sporting events throughout the world and other instances have been associated with crowd panic.

Hog-tying

A number of cases have been described where an individual, often a person requiring restraint during arrest, has been tied up, usually with the wrists and ankles tied together behind the back. If such an individual is left lying face down on a hard surface then the restricted movements induced by the ligatures together with possible reduced chest movement caused by an apparent splinting effect produced by contact with the floor, may result in hypoxia and sudden death. This method of restraint is known as hog-tying. Variations of this form of death may be associated with sexual asphyxiation described below.

In North America, hog-tying is a not infrequent method of restraining prisoners during the process of arrest. Sudden death has been reported in some of these individuals. In some instances, however, the victims have been intoxicated with alcohol and or using cocaine. It may be that chemical suppression of respiration is a factor in such deaths. These cases often show evidence of injury associated with the ligatures but no other evidence of asphyxiation.

Pickwickian syndrome

A form of chronic asphyxiation associated with respiratory failure is the Pickwickian syndrome. This condition, named after the Pickwickian fat boy who kept falling asleep, is associated with extreme obesity where, as a result of obesity and excess body weight, diaphragmatic and intercostal movement is significantly reduced. The victims become chronically hypoxic, and develop high carbon dioxide levels with associated narcosis.

Death by crucifixion

This form of punishment was apparently introduced to the Romans by the Phoenicians. The victim was suspended from a cross in such a fashion that in order to breathe, he was required to raise his body by pressing down on his feet which were either nailed or supported by part of the mechanism. The victim rapidly became exhausted and death occurred when the muscles of respiration failed and right heart failure occurred. In the case of inverted crucifixion as in the death of St Peter, respiratory paralysis would be induced by the weight of the abdominal viscera.

Cases of accidental inverted suspension causing death have been documented and in all of these cases, death is as a result of respiratory exhaustion. With the exception of the circumstances of death, these cases are indistinguishable from those of postural asphyxia.

Mechanical Obstruction by Neck Compression

Strangulation

Strangulation, sometimes referred to as throttling, is divided into two main types: manual and ligature. Not infrequently however, in the homicidal situation, strangulation may be as a result of the combination of manual and ligature compression of the neck. Strangulation is usually defined as the application of a compressing force to the neck where the force acting on the neck is other than that of the weight of the body. This distinguishes it from deaths by hanging described below.

Strangulation, both manual and ligature, probably represents the most common cause of homicidal asphyxiation. Manual and ligature strangulation are both associated with external and internal neck injuries, which in some cases may be minimal, but in most cases are easily seen. Petechial hemorrhages are also a characteristic, although not diagnostic, finding in such deaths. When such petechiae occur in cases of strangulation, they are almost always distributed above the level of neck obstruction and distal to the heart. Although such petechiae are characteristically seen in strangulation, and their distribution should be noted in detail, it is most important to recognize that petechiae are a relatively common nonspecific

autopsy finding. They occur in a wide variety of non-asphyxial states and may even be a postmortem phenomenon associated with the postmortem posture of the body. In strangulation, there may be partial or complete obstruction and compression of all of the principal anatomical structures in the neck, including the veins, arteries, airway and nerves.

Occlusion of the neck veins is almost solely responsible for the classic signs of congestion, cyanosis and petechiae above the level of neck constriction. Pressures as low as 2 kg will compress the jugular veins and obstruct venous return to the heart. This can cause a rapid rise in venous pressure in the head with the associated rupture of small venules.

Airway obstruction Strangulation may be associated with either partial or complete obstruction to the airway by direct compression of either the larynx or trachea. Obstruction at the level of the pharynx may be produced by manual elevation of the larynx so that air entry is blocked by the root of the tongue being pressed against the palate. Compression of the larynx may fracture the pharyngeal cartilages and hyoid bone, although the number distribution and nature of these fractures will depend on the age of the victim and degree of calcification and fusion of these structures. Collapse of the larynx associated with these fractures may, of itself, cause airflow obstruction. Various experiments have shown that forces of approximately 15 kg will close the trachea. This is substantially more than is required to occlude the blood vessels.

Arterial compression Compression of the arteries in the neck is less common than venous occlusion since the carotid arteries are largely obscured by the sternomastoid muscles. Bilateral occlusion of the carotids may cause rapid loss of consciousness as the arterial blood supply to the brain is reduced and it appears that vertebral artery circulation alone is insufficient to maintain cortical function. It seems unlikely, therefore, that occlusion of the carotid arteries alone would be sufficient to cause death. Occlusion of the vertebral arteries by neck compression appears to be almost impossible because of the anatomical location of these vessels. The general opinion is that permanent brain damage is very unlikely if arterial occlusion persists for less than 4–5 min and it has been reported that even in normothermic conditions total recovery has occurred after 9–14 min of vascular occlusion.

Neurological effects (vagal inhibition) Direct pressure or manipulation of the baroreceptors situated in the carotid sinuses (carotid bodies) can result in reflex bradycardia or total cardiac arrest. This mechanism acts through a vagal nerve reflex arc arising in the complex nerve endings of the carotid sinus and returning via the brainstem and the vagus nerve to exert its slowing effect on the heart.

Susceptibility to vagal cardiac arrest appears to be variable and it has been suggested that fear, apprehension and possibly struggling may heighten sensitivity to this vagal mechanism by the release of catecholamines. The amount of force that must be applied to the carotid sinuses and the duration of the application of such force is variable and unpredictable. There are many well-documented examples of sudden impacts to the neck causing cardiac arrest and other cases where the application of simple neck massage has induced death.

If cardiac arrest occurs quickly during compression of the neck, then petechial hemorrhages which are often associated with relatively long periods of neck compression lasting more than 2 min, may be minimal. It seems likely that a vagal cardiac arrest component of death in strangulation cases is much more likely in manual strangulation than ligature strangulation because of the ability of the fingers to compress and massage the carotid sinus area of the carotid arteries. Most forensic pathologists will have experienced cases where this mechanism of apparently sudden death in neck compression cases has occurred, and it is the opinion of this writer that vagal inhibition is an important component of death or incapacitation in many cases of manual strangulation.

A frequently asked question of forensic pathologists is how long must neck compression be applied before an individual is rendered unconscious or dies. Because of the complexity of the mechanisms of incapacitation, loss of consciousness and death in strangulation, there is no simple answer to this question. It would appear that for petechial hemorrhages to occur, the compression must last for at least 30 s. However, although it is probably correct, in most cases, to suggest that the development of petechial hemorrhages is an indication of a sufficiently long-time compression to induce unconsciousness, it is not possible to determine this period of time with any degree of accuracy.

Manual strangulation

The autopsy findings in manual strangulation include the signs of injury to the neck, both external and internal, and the signs of the mechanism of death.

Manual strangulation is often associated with external bruising of the neck which may be more prominent on one side than the other. These bruises are usually caused by compression by the pads of the

fingers or thumb. They tend to be at the level of the thyroid cartilages and are on either side of the front of the neck. Internal examination of the neck structures requires layer-by-layer dissection of the neck tissues *in situ*. Only after such careful dissection is it possible to identify bruising at various levels within the neck tissues. In some cases, the evidence of external bruising may be minimal but careful dissection of all the anterior and posterior neck structures will reveal evidence of deep bruising.

Abrasions may be found on the skin usually at the front of the neck. These abrasions may be characteristically linear or semilunar in form and are usually caused by fingernails. In many instances it is the fingernails of the victim trying to release the compressing hand which produce the injuries. It is probably more frequent to have fingernail marks produced by the victim than by the fingernails of the assailant.

Prominence of bruising on one side of the neck may suggest that the pressure has been caused by a thumb, rather than a group of fingers, but this is not necessarily correct. As a general rule it is wrong to use asymmetrical grouping of bruising in the neck as evidence of right or left handedness. Such grouping of bruises may be entirely fortuitous and may be determined as much by the relative positions of victim and assailant, as the position of the hand of the assailant. In some cases, patterned injuries may be present on the front of the neck when a textured garment lies between the compressing hand and the skin.

In cases of bodies undergoing early decomposition, particularly if the bodies have been lying in a prone position, there may be hemorrhage into the tissues of the neck and between the layers of neck muscle. These bruises are purely artifactual associated with vascular decomposition and gravity-dependent postmortem bleeding. In manual strangulation, the greater cornuae (lateral horns of the hyoid bone) may be fractured. However, in children and young adults, incomplete ossification and union of the components of the hyoid bone may make it almost impossible for fractures to occur. Because of the significance of such fractures, it is most important that removal of the hyoid bone and pharyngeal cartilages at the time of postmortem should be carried out with the utmost care.

Radiological examination may fail to reveal any fractures. Careful dissection of the cartilaginous structures is usually required to identify the fractures. The presence of bruising and hemorrhage associated with fractures to the larynx and hyoid bone may be considered evidence of the antemortem nature of the injuries. In the absence of such bruising and hemorrhage the possibility that the fractures are post mortem must be considered.

Fractures of the hyoid bone and laryngeal cartilages can occur as a result of direct blows to the neck and, therefore, in the absence of other evidence of strangulation, are not necessarily diagnostic of strangulation.

Because of the importance of identifying evidence of vagal inhibition as a possible factor in death, dissection and microscopic examination of the region of the carotid artery bifurcation including the carotid sinus is important and should be a routine part of the examination in alleged strangulation cases.

Ligature strangulation

Pressure on the neck by a constricting ligature, sometimes called ‘garroting’, frequently leaves a prominent grooved and abraded ligature mark around the neck. Such ligature marks, however, are not necessarily circumferential. The ligature mark although invariably horizontal may be interrupted by the positioning of clothing or long hair. Cords, wires, ropes and belts can be used as ligatures. Occasionally soft fabrics such as scarves and towels may leave little evidence of a ligature mark. In many cases, the ligature mark is not a simple linear mark but may be complex with evidence of criss-crossing of the ligature. The complexity of the mark may in part be caused by this overlapping of the ligature but may also be caused by either movement of the ligature on the neck or a double or repeated application of the ligature.

Although the majority of cases of ligature strangulation are homicidal in nature, suicidal ligature strangulation can occur. However, because, in the process of strangulation, the victim will become unconscious, self-strangulation requires a ligature which is either elastic and tightens itself around the neck, or some form of twisting device, such as a Spanish windlass which will prevent the ligature loosening.

Injuries to the neck musculature tend to be less marked in ligature strangulation than manual strangulation and may be entirely confined to the area immediately beneath the ligature.

Fractures of the hyoid bone and laryngeal cartilages are much less frequently seen in ligature strangulation than in manual strangulation. However, dissection and microscopic examination of the larynx may reveal evidence of internal hemorrhage.

Ligature strangulation is more often associated with the classic signs of asphyxia than manual strangulation unless reflex vagal inhibition with a cardiac arrest occurs and thus limits the development of these signs. Congestion and cyanosis above the level of the

ligature may be prominent. The veins are completely obstructed in ligature strangulation but complete occlusion of the arterial system is relatively rare.

Neck holds and arm locks

Neck holds are taught by law enforcement agencies as a means of subduing suspects resisting arrest or to control prisoners who are combative and unmanageable. There are two basic types of neck holds. The 'carotid sleeper hold' often portrayed in professional wrestling competitions, and the 'bar arm hold' or 'choke hold'.

Generally speaking, the choke hold, which involves pressure by the forearm of the assailant against the front of the neck of the victim, may cause collapse of the airway and serious injury to the structures of the neck. The carotid sleeper hold preserves the airway and is designed to compress the common carotid arteries producing transient cerebral ischemia and unconsciousness. The dangers associated with these holds have been researched and well documented. It is well recognized that deaths may occur during the application of these holds and during an arrest or controlling situation, where the victim is struggling, it may be almost impossible to apply a sleeper hold correctly. There is also some evidence that victims intoxicated with cocaine may be at increased risk of cardiac arrest and sudden death in such restraint situations.

Hanging

Hanging is a form of ligature strangulation in which the force applied to the neck is derived from the gravitational weight of the body or part of the body. In hanging, suspension of the body does not have to be complete. The forces required to compress the various neck structures and the clinical effects of experimental hanging have been fully documented.

Hangings may be accidental, suicidal, or homicidal. Accidental hanging is relatively uncommon but there are many cases recorded where adults and children have been accidentally hanged by being suspended after becoming entangled in ropes or clothing.

The commonest from of hanging is suicidal hanging. The degree of suspension is very variable and the victim's feet are often resting on the ground and in some cases, the victim may be sitting on the ground with the force applied to the ligature being only that of the upper part of the body.

Homicidal hanging, with the relatively rare exception of lynching, is very rare and is usually associated with other injuries indicating a struggle prior to the hanging process. In homicidal hanging there are often other detectable reasons for incapacitation such as intoxication or head injury. In a few cases of homicidal strangulation, the body is suspended after death to make the case appear to be one of suicidal hanging.

The mark on the neck in hanging can almost always be distinguished from ligature strangulation. Because the force applied to the ligature is that of the weight or part of the weight of the body, the ligature is rarely horizontally positioned around the neck. It is usually directed upwards indicating the point of suspension and producing a ligature mark at the point of suspension, which may resemble an inverted V. A hanging mark often does not completely encircle the neck. Depending on the type of knot used and the tightness of the ligature, it is possible for relatively complex marks to be produced on the neck as the ligature slips, stretches or tightens around the neck structures.

In hanging cases, the surface texture of the ligature mark is frequently reproduced as a parchmented, grooved patterned abrasion characteristic of the particular ligature. The spiral weave or woven surface of a rope or cord may be accurately reproduced on the skin.

The autopsy findings in hanging frequently involve only the ligature mark on the outside of the neck. There may be no internal bruising detected, even after careful dissection and petechial hemorrhages in cases of complete suspensions are relatively rare. In hanging cases, with petechial hemorrhages distributed above the level of the ligature, the suspension has usually been incomplete. Similar effects can also be produced if the ligature breaks and the tension of the ligature is relieved. Dissection of the carotid arteries may show occasional linear traction (stretch tears) of the intima and fractures of the vertebrae do not occur unless there is a significant drop such as in judicial hanging. Even in judicial hanging fractures are relatively rare although bruising and tearing of the intervertebral ligaments may be seen.

Sexual Asphyxias

Auto-erotic or masochistic practices may involve compression of the neck inducing partial asphyxiation, alleged to heighten sexual pleasure and induce orgasm. Sexual asphyxias occur almost exclusively in men with only a few reports in the literature involving women. In the majority of deaths associated with sexual asphyxiation, there has been complete or partial suspension of the body with complex systems of ligature application, usually involving at least

compression of the neck and obstruction of respiration. These deaths are usually accidental because the fail-safe mechanisms built into the routine of the practitioner have failed.

Characteristic findings in sexual asphyxias may include padding of the ligatures to prevent permanent injury, the presence of erotic or pornographic materials at the scene of death or evidence of prior ligature-related activities including grooves or damage to suspension points. The victims often arrange mirrors or cameras so that they can either witness or record their activities. Many of these activities are associated with a variety of fetish activities, transvestitism, cross-dressing and other evidence of sexual activity including binding or piercing of the genitalia and breasts.

See also: **Causes of Death**: Postmortem Changes. **Toxicology**: Inhalants. **Autoerotic Death**.

Further Reading

DiMaio DJ and Di Maio VJM (1989) *Forensic Pathology*. Boca Raton: CRC Press.

Gordon I, Shapiro HA and Berson SD (1988) *Forensic Medicine – A Guide To Principles*, 3rd edn. Edinburgh: Churchill Livingstone.

Humphry D (1991) *Final Exit*. Eugene: The Hemlock Society.

Jaffe FA (1994) Petechial hemorrhages. A review of pathogenesis. *American Journal of Forensic Medicine and Pathology* **15**:203–207.

Jaffe FA (1997) Pathogenicity of carbon monoxide. *American Journal of Forensic Medicine and Pathology* **18**:406–410.

Knight B (1996) *Forensic Pathology*, 2nd edn. London: Arnold.

Polson CJ, Gee DJ and Knight B (1985) *The Essentials of Forensic Medicine*, 4th edn. Oxford: Pergamon Press.

Raven KP, Reay DT and Harruff RC (1999) Artifactual injuries of the larynx produced by resuscitative intubation. *American Journal of Forensic Medicine and Pathology* **20**:31–36.

Reay DT and Eisele JW (1982) Death from law enforcement neck holds. *American Journal of Forensic Medicine and Pathology* **3**:253–258.

Reay DT and Howard JD (1999) Restraint position and positional asphyxia. *American Journal of Forensic Medicine and Pathology* **20**:300–301.

Tardieu A (1855) Memoire sur la mort par suffocation. *Ann Hyg Pub Med Legale* (Series 11) **vi**:371–382 (cited by Knight B (1996)).

Thomas JE (1969) Hyperactive carotid sinus reflex and carotid sinus syncope. *Mayo Clinical Proceedings* **44**:127–139.

Blunt Injury

S Pollak, University of Freiburg/Br, Freiburg, Germany
P J Saukko, University of Turku, Turku, Finland

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0427

Introduction

The term 'blunt trauma' can be defined as damage to the body due to mechanical force applied either by the impact of a moving blunt object or by movement of the body against a hard surface, both mechanisms resulting in the transfer of kinetic energy high enough to produce an injury (mainly by compression, traction, torsion and shear stresses). Blunt force injuries occur in criminal assaults (e.g. blow with a blunt-edged instrument, punch, kick), in physical child abuse, in traffic and industrial accidents, in suicides (jump from a height) and in accidental falls brought about by the victim himself.

Blunt Injuries to the Integument

Abrasions

Abrasions are superficial injuries to the skin characterized by a traumatic removal, detachment or destruction of the epidermis, mostly caused by friction. In so-called tangential or brush abrasions a lateral rubbing action scrapes off the superficial layers of the skin (for example from the body's sliding across a rough surface) and leaves a denuded corium which is initially covered with serosanguineous fluid. In fresh grazes the direction of impact can often be determined by abraded epidermal shreds which remain attached to the end of the scrape. At a later time the tissue fluid dries out and forms a brownish scab. If the lesion does not reach the dermis, it heals within several days without scarring. Infliction just before or after death results in a leathery ('parchment-like') appearance with yellowish-brown discoloration (**Fig. 1**).

Another type of abrasion is caused by a vertical impact to the skin (so-called pressure or crushing abrasion). In such cases the injuring object may be reflected by the shape of the skin injury, so that the patterned abrasion can be regarded as an imprint of the causative object.

Contusions

Contusions or bruises are extravasations of blood within the soft tissues originating from ruptured vessels as a result of blunt trauma. In this context,

Figure 1 Tangential abrasion confined to the upper layers of the skin. The shreds of epidermis indicate the direction of impact. The central parts of the nuded corium show a parchment-like discoloration from drying.

only the contusions that are visible externally are considered. Textbooks usually differentiate between intradermal and subcutaneous bruises. In the first-mentioned category the hemorrhage is located directly under the epidermis, i.e. in the corium. This kind of superficial hematoma is usually sharply defined and red, whereas the more common bruises of the deeper subcutaneous layer have blurred edges and, at least initially, a bluish-purple color (**Fig. 2**).

Intradermal bruises may reflect the surface configuration of the impacting object (Fig. 2). The skin which is squeezed into grooves will show intradermal bleeding, whereas the areas exposed to the elevated parts remain pale. Especially in falls from a height, the texture of the clothing may produce a pattern of intradermal bruises corresponding to the weaving structure. Patterned extravasations of this type are also seen in tire tread marks, if an individual is run over by a wheel, and in bruises from vertical stamping with ribbed soles.

Figure 2 Patterned contusions from kicking to the thigh (partly in combination with curved abrasions corresponding to the front part of the perpetrator's shoe).

Subcutaneous bruises are usually nonpatterned. Nevertheless there may be bruising of special significance. If the body is struck by a stick, a broom-handle, a pool cue, a rod or another elongated instrument, every blow leaves a double 'tramline' bruise consisting of two parallel linear hematomas with an undamaged zone in between. Victims of blunt force violence often sustain contusions from self-defense, typically located on the ulnar aspects of the forearms and on the back of the hands. The upper arms may show groups of roundish bruises from fingertip pressure in cases of vigorous gripping. A periorbital hematoma ('black eye') is either induced by direct impact (e.g. punch or kick) or indirectly (due to seepage of blood from a fractured orbital roof, a fractured nasal bone or from a neighboring scalp injury of the forehead).

In general bruises are regarded as a sign of vitality indicating that the contusion was inflicted prior to death. During life the blood from ruptured vessels is forced into the soft tissue by active extravasation. Nevertheless, to a limited extent postmortem formation of contusions is possible due to passive ooze of blood. In surviving victims, a deep bruise may not become apparent on the skin until several hours or even days later because of the slow percolation of free blood from the original site to superficial tissue layers.

In a living person the contusion undergoes a temporal series of color changes. Initially most subcutaneous bruises appear purple–blue. As the hematoma resolves during the healing process, the hemoglobin released from the red blood cells is chemically degraded into other pigments such as hemosiderin, biliverdin and bilirubin. The color changes – usually in the course of several days – to green and yellow before it finally disappears. However, the rate of change is quite variable and depends on numerous factors, above all the extent of the bruise.

The size of an intradermal or subcutaneous hematoma is not always indicative of the intensity of the force applied to the affected area. Elderly people or patients suffering from bleeding diathesis may get bruises from slight knocks or for other minor reasons. On the other hand, absence of an externally visible injury does not necessarily mean that there was no relevant trauma. Subcutaneous bruises of surviving victims are often followed by gravity shifting of the hemorrhage leading to a secondary downward movement of the hematoma.

A special type of blunt injury to the soft tissues is frequently seen in pedestrians who have been struck or run over by motor vehicles. Both the skin

and the subcutaneous layer may be avulsed from the underlying fascia or bones by shearing forces so that a blood-filled pocket is formed, typically in combination with crush damage to the adjoining fatty tissue.

Lacerations

Lacerations are tears of the skin or of internal organs (see below). They may be caused by blows from blunt objects (such as a hammer, a whipped pistol, a rod, a toe-cap of heavy footwear or a fist); other lacerations are produced by impact from vehicles or by a fall to the ground. Lacerations occur most commonly in body regions where the integument directly overlies a firm bony base acting as support (scalp, face, back of the hand, shins). When the force acts on the skin, the subcutaneous tissue is squeezed between the injuring object and the bony platform so that the integument is compressed and crushed until it tears and splits sideways (**Figs 3** and **4**).

Lacerations are characterized by abraded, bruised and crushed wound margins. The edges of the tears are typically irregular and ragged with bridging tissue strands (vessels, nerves and fibers) running from side to side. The wound slits may be linear (especially in blows with a narrow, edged instrument), Y-shaped or star-like. If the impacting object hits the skin at an oblique angle, one of the edges will be ripped away resulting in unilateral undermining (undercutting, avulsion) which indicates the direction of the force. Sometimes foreign material from the causative instrument/surface is deposited in the depth of the wound slit. The abrasion surrounding the tear may correspond to the shape and dimensions of the impacting blunt-surfaced instrument (Fig. 4) or – in case of a fall to the ground – the area of contact.

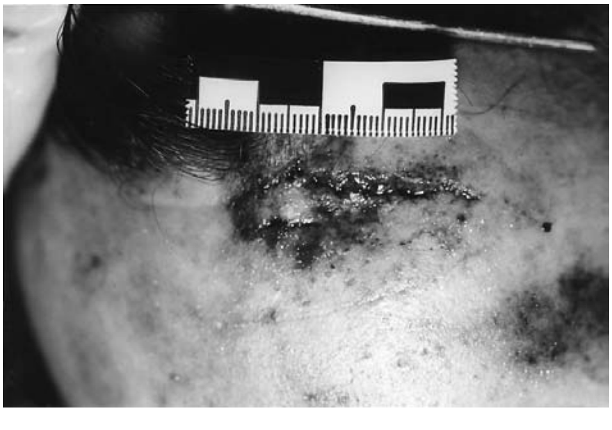


Figure 3 Laceration to the forehead. Note the irregular wound slit with abraded and contused margins.

Figure 4 Parietal region of a man whose scalp has been shaved before autopsy, showing two stellate lacerations from hammer blows to the head. The abrasion which belongs to the upper injury approximately reproduces the shape of the rectangular hammer face.

Head Injuries

The head is a common target in assaults with blunt objects; other frequent causes of head injuries are traffic accidents, falls from a height and falls from a standing position. The area of impact usually reveals injuries of the scalp or the facial skin, but it has to be stressed that severe and even lethal traumatization is not necessarily associated with scalp bruising, marked swelling, excoriation and/or laceration. There may be no externally visible signs, especially in skin areas covered with hair and in cases of a fall onto a flat surface. An impact site on the vertex suggests that the head sustained a blow whereas in falls from standing positions the scalp injuries are expected at the level of the brim of the hat.

Skull fractures

These may involve the cranial vault, the base of the skull and the facial skeleton. Though the presence of a skull fracture indicates severe traumatization, the fracture itself rarely threatens the victim's life. There are several *types of* skull fractures to be distinguished.

Single or multiple *linear fractures* are caused either by a blow with an object with a broad flat surface area or by a fall on the head so that the skull is deformed (flattening/indenting at the point of impact and outward bending/bulging in the periphery). The

fracture lines originate where the bone is bent outwards and therefore is exposed to traction forces exceeding the limits of the bone's elasticity; from these extruded parts of the skull the fractures extend towards the area of impact, but also in the opposite direction. For this reason, either of the ends is often in congruity with the impact injury of the scalp. Several fracture lines may radiate outward from a central point of impact (**Fig. 5**) where the skull is often depressed and/or shattered to pieces forming a *spider's web* or mosaic pattern consisting of circular and radiating linear fractures. The sequence of skull injuries may be determined according to Puppe's rule: a later fracture does not cross a pre-existing fracture line but terminates when reaching an earlier one.

Before fusion of the cranial sutures (i.e. in children and young adults), a fracture may travel along the seam resulting in diastasis (*diastatic fractures*). If a gaping fracture runs from one side of the cranial base to the other (mostly after lateral impact or side-to-side compression), this transverse type is called a *hinge fracture* (**Fig. 6**) because of the independent movement of the front and rear halves of the skull base.

Longitudinal fractures of the base of the skull frequently occur due to a fall on the occiput; in such instances the linear fractures typically run through the posterior fossa either ending near the foramen magnum (**Fig. 7**) or extending to the floor of the middle and anterior fossa. On the other hand, longitudinal fractures of the base can also be produced by impaction of the frontal region.

Depending on its course and location, a *base fracture* may be followed by several *clinical signs*: bleeding from the ear (in fractures of the temporal bone with concomitant hematotympanon and rupture of the eardrum); bleeding from the nose and mouth (in fractures involving paranasal sinuses which provide a communication with the nasopharynx); periorbital hematoma (from fractures of the orbital roofs); leakage of cerebrospinal fluid coming out of the nose or the ear (if the dura is injured along the fracture); bacterial infection of the meninges (by spread from the nasal cavity, the paranasal sinuses and the middle ear, especially when the fracture is accompanied by a tear of the dura).

Figure 6 Transverse (side-to-side) 'hinge fracture' dividing the base of the skull into a front and rear half.

Some special types of skull fractures can only be mentioned briefly. A *ring fracture* is located in the posterior fossa and encircles the foramen magnum. It mostly occurs in a fall from a height onto the victim's feet or buttocks so that the cervical spine is driven into the skull.

Bone impressions and *depressed fractures* are always localized at the point of impact where the head is struck with an object having a relatively small surface area such as a hammer or a protruding corner of a piece of furniture (**Fig. 8**). The outline of a clean-cut defect in the outer table may reproduce the shape and size of a sharp-edged instrument. If only limited force is applied, the depressed fracture can be restricted either to the outer or, less often, to the inner table of the skull (the latter with inward displacement of the bone fragments). A depressed fracture from a blow which struck the skull cap at an angle may be concentrically terraced. *Hole fractures* from bullets perforating a flat bone of the skull are mostly roundish and clean-cut at the site of entrance, but beveled out in a crater-like manner at the exit site.

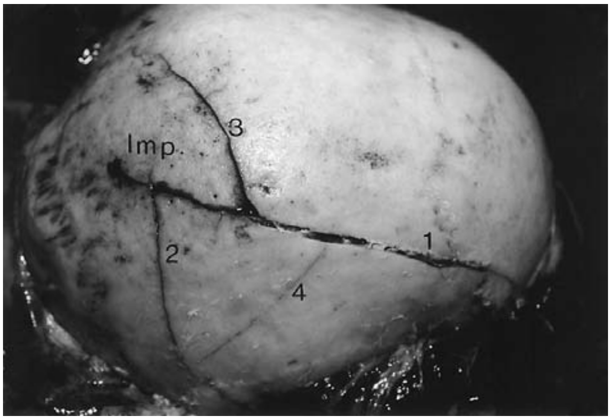


Figure 5 Right half of a cranial vault with two fractures (1, 2) radiating outward from the area of impact (Imp.) and in addition two incomplete circular fractures (3, 4).

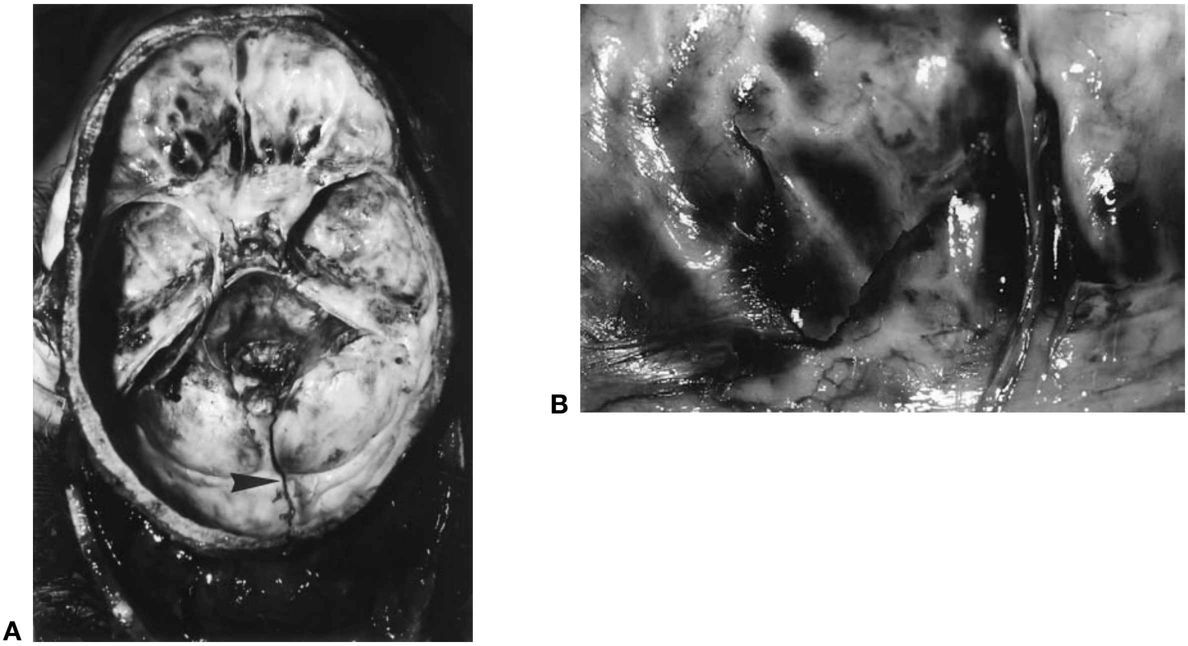


Figure 7 (**A**) Base of the skull with linear fracture (arrow) in the posterior fossa extending downwards into the foramen magnum, sustained in a fall on the occiput. Note the presence of secondary fractures in the anterior fossa floor. (**B**) Close-up view of the fractured left orbital roof with dark coloration from underlying blood.

Blunt force applied to the occiput, mostly as a consequence of a fall on the back of the head, frequently causes independent fractures of the anterior cranial fossa such as cracks of the thin orbital roofs (*secondary fractures* at the site of the contrecoup).

Intracranial hemorrhages

A space-occupying bleeding into the brain membranes is followed by local displacement of the brain and raised intracranial pressure with concomitant flattening of the cerebral hemispheres.

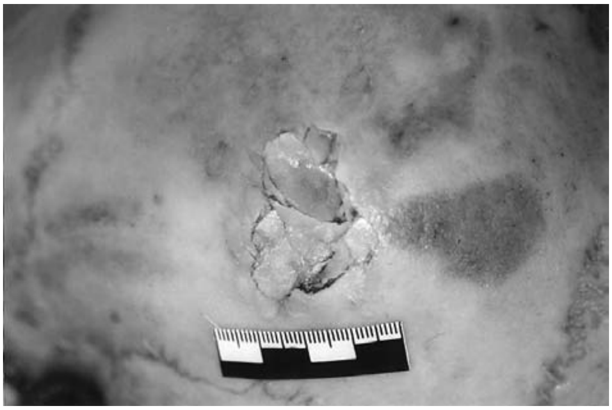


Figure 8 A depressed fracture of the skull cap from a blow with a heavy monkey-wrench.

Intracranial hematomas as well as traumatic brain swelling, which often accompanies head injuries, may result in transtentorial (uncal) herniation (in cases of supratentorial mass lesion) and/or herniation of the cerebellar tonsils which are forced into the foramen magnum leading to compression of the brainstem with secondary damage and failure of the medullary respiratory centers.

From the clinical and forensic point of view, the possible occurrence of a so-called 'lucid' or 'latent' interval has to be mentioned. After initial unconsciousness (due to cerebral concussion) there may be a symptomless period of several hours or even days before the victim becomes comatous again because of the increased hemorrhage and the consequently raised intracranial pressure.

Epidural (extradural) hemorrhages are located between the skull and the underlying dura mater which is stripped from the bone by bleeding from a torn vessel (**Fig. 9**). Epidural hematomas have a typical disc- or lens-shaped appearance. The most common site is the temporal and the adjacent parietal region where the branches of the middle meningeal artery are easily lacerated in the course of a transecting fracture line. Since the well-adherent dura has to be avulsed from the bone, epidural hematomas more frequently originate from arterial bleeding than from venous bleeding (e.g. due to a torn dural sinus). In the great majority an extradural hemorrhage is associated with a cranial fracture.

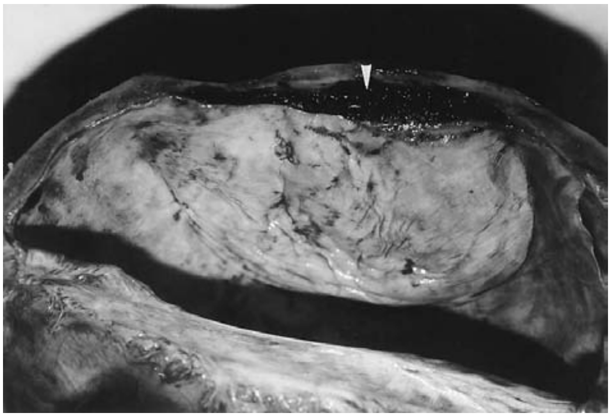


Figure 9 Extradural (epidural) hemorrhage in the parieto-temporal area. The hematoma (arrow) is located between the inner surface of the skull and the detached dura.

Subdural hematomas are intracranial bleedings located beneath the dura mater and above the arachnoid (**Fig. 10**). Most often the hemorrhage arises from the tearing of overstretched bridging veins that traverse the subdural space between the surface of the cerebral hemispheres and the superior sagittal sinus. Other possible sources of subdural bleeding are injuries to venous sinuses or to the cerebral parenchyma (such as cerebral contusions with concomitant laceration of the arachnoid). The subdural hemorrhage usually covers one cerebral hemisphere in a cap-like manner from the parasagittal area via the lateral surface down to the basal fossas; on a horizontal section it appears as a sickle-shaped accumulation of blood. In contrast to epidural hematomas, subdural hemorrhages are often not associated with skull fractures; additional damage to the brain tissue may also be absent. A high percentage of subdural bleedings are caused by acceleration or deceleration of the head, for instance in falls when the head impacts a hard surface, but also in traffic accidents and physical child abuse (battered child and shaken baby syndrome). Apart from acute and subacute subdural hemorrhages there are prolonged cases of hematoma formation and organization, mainly in elderly people and sometimes without a history of previous traumatization. Such chronic subdural hematomas typically consist of brown and gelatinous blood accumulations adherent to the meninges and sometimes covered with a tough membrane.

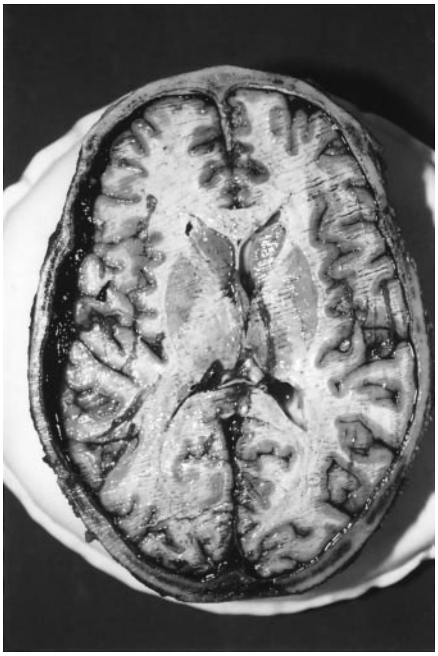


Figure 10 Acute subdural hemorrhage. The unilateral space-occupying lesion has slightly shifted the midline of the brain to the opposite side.

Traumatic subarachnoid bleeding may result from damage to the cortex such as brain contusion (e.g. contrecoup lesions), from penetrating injuries to the brain and as a consequence of vessel tears within the subarachnoid space. An extensive hemorrhage on the ventral surface and around the brainstem may arise from a laceration of an artery belonging to the circle of Willis or from another great vessel (such as a torn basilar and vertebral artery).

Cerebral injuries

'*Concussion* of the brain' is a clinical diagnosis which means a disorder of cerebral function following immediately upon a (blunt) head injury. It is usually characterized by a transient loss of consciousness (initial coma) with subsequent amnesia from the actual moment of trauma; it is often combined with retrograde amnesia and vegetative signs such as nausea and vomiting. In mere concussions the unconsciousness lasts only for a relatively short time (less than 1 h) and the brain tissue does not show any evidence of structural damage. Nevertheless even a simple cerebral concussion may be followed by the victim's death, if the head trauma is joined by interfering mechanisms (for instance drowning or aspiration of gastric contents during unconsciousness).

Cerebral contusions are traumatic lesions of the brain frequently seen in the cortex and sometimes extending into the underlying white matter (**Fig. 11**). Fresh contusion hemorrhages are mostly located on the crests of the gyri and composed of grouped streak-like or punctate blood extravasations. The cortical lesions are often covered with subarachnoid bleeding. In contrast to cerebral contusions, the term 'laceration' means a major destruction of the

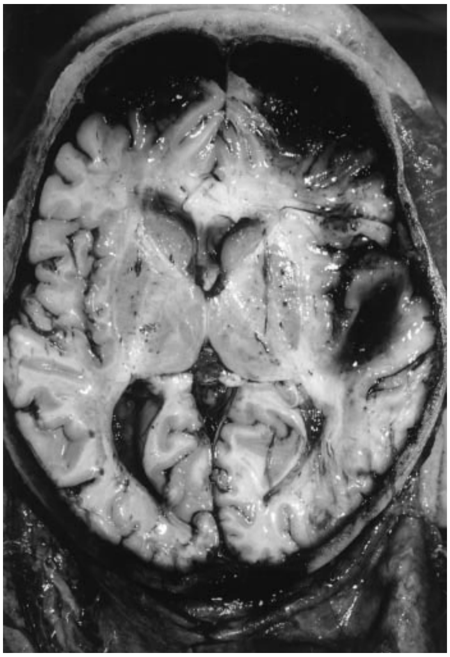


Figure 11 Contrecoup contusions of the frontal poles opposite to the point of impact (fall on the back of the head) with concomitant subarachnoid hematoma; traumatic intracerebral (subcortical) hematoma in the white matter of the right temporal lobe (in close vicinity to cerebral contusions in the overlying cortex).

anatomical context (for instance mechanical separation of the tissue due to bone fragments or penetrating bullets). In case of survival, the contusion hemorrhages are reabsorbed and assume a yellowish-brown appearance with softening and finally liquefaction of the affected areas.

Due to the injuring mechanism, most cerebral contusions occur in brain regions that are directly opposite to the point of impact. This contrecoup type of contusion is classically caused by a fall on the occiput, when the moving head is suddenly decelerated with the consequence that the inlying brain is damaged due to inertia. In falls on the back of the head, the contrecoup areas of the brain (poles and undersurfaces of the frontal and temporal lobes) are subjected to an ultrashort negative pressure ('cavitation') resulting in vessel ruptures and cortical hemorrhages (**Figs 12** and **13**). On the other hand, the so-called coup contusions arise at the area of impact due to the local deformation and compression of the brain. Even severe coup and contrecoup injuries are not necessarily associated with skull fractures. In victims with both coup and contrecoup lesions, the degree of contrecoup damage is usually more marked. Fracture contusions are localized in topographical correspondence to fracture lines and/or depressed fractures.

Figure 12 Contrecoup contusions on the undersurface of the frontal lobes (mostly located at the crests of the convolutions) with slight subarachnoid hemorrhage.

Diffuse axonal injury (DAI) is considered a consequence of shear and tensile strains from sudden acceleration/deceleration or rotational movements of the head. Overstretching of the nerve fibers in the white matter leads to axonal injury varying from temporary dysfunction to anatomical transection, the latter being followed by microscopically visible club-shaped retraction-balls on the axons. The sites of predilection include the corpus callosum, the parasagittal white matter, the superior peduncles and the upper brainstem. In the course of the repair process, microglial cells proliferate in the areas of axon damage. In victims of substantial head injuries, especially after traffic accidents, diffuse axonal injury may be responsible for prolonged coma and a fatal outcome even in the absence of an intracranial mass lesion.

Cerebral edema is a frequent finding in significant head injuries. The formation of edema is due to an

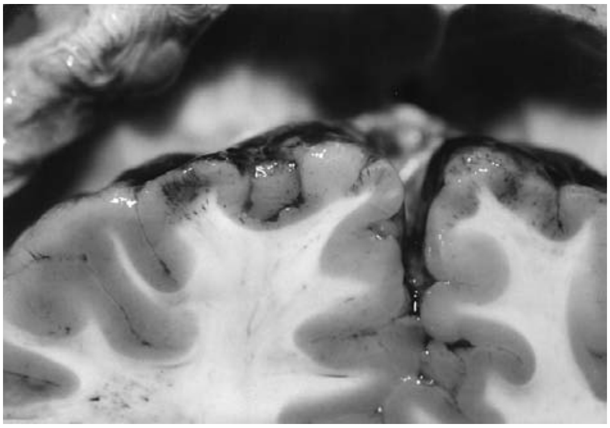


Figure 13 Close-up view of the frontal lobes with characteristic streak-like, densely arranged contusion hemorrhages in the cortex.

increase in the fluid content of the brain, predominantly in the white matter. Post-traumatic edema may be generalized (diffuse) or related to focal tissue damage (e.g. adjacent to an area of cerebral contusion or laceration). At autopsy the weight of the brain is increased, the gyri are pale and flattened with shallow sulci in between. From the pathogenetic point of view, edema is attributed to a heightened vascular permeability which in turn may be worsened by additional hypoxia.

As with space-occupying lesions such as subdural or epidural hematomas, cerebral edema is a common cause of raised intracranial pressure. The enlarged volume of the edematous brain results in a displacement of cerebral tissue downwards through the midbrain opening resulting in grooving of the unci and/or hippocampal herniation. Expansion of the subtentorial brain leads to herniation of the cerebellar tonsils which are forced into the foramen magnum. Herniation with concomitant compression of the brainstem may be followed by secondary hemorrhages (localized in the midbrain and pons) and finally by lethal dysfunction of the vital centers.

Injuries of the Chest

Nonpenetrating blunt force may damage the thoracic wall and/or the chest organs. *Rib fractures* are caused either by direct or by indirect violence. In the first case a localized force is applied and the underlying ribs are broken in the contact area; the other (indirect) type of rib fracture occurs away from the impact, mainly due to compression of the chest.

Rib fractures are frequently associated with *complications* that may be dangerous or even life-threatening.

- If a victim sustains numerous fractures, the rib cage loses its rigidity so that the injured section of the chest wall will not participate in the expansion of the thorax during inspiration with the result of paradoxical respiration (*flail chest*) and concomitant hypoxia.
- Sharp, pointed ends of the rib fragments may penetrate the pleura and lacerate the lung and/or the intercostal blood vessels with consecutive bleeding into the chest cavity (*hemothorax*).
- A leak in the visceral pleura permits air to enter the pleural cavity (*pneumothorax*) so that the lung collapses, if it is not fixed to the chest wall by pre-existing pleural adhesions. A valve-like leakage in the pleura leads to a so-called tension-pneumothorax caused by an increasing pressure of trapped air in the pleural cavity and followed by a complete collapse of the affected lung and a shift of the mediastinum to the opposite side.
- The presence of air bubbles in the subcutis or in the mediastinum (*subcutaneous/mediastinal emphysema*) may derive from injuries of the trachea, the bronchi, the thoracic wall or the lungs by air entering the adjacent soft tissues.

Blunt-force injuries to the *lung* are mainly encountered as contusions or lacerations. A contusion is typically caused by a substantial impact on the chest with consecutive inward bending of the thoracic cage. In young victims contusions are not necessarily accompanied by fractures of the ribs or of the sternum because of the high pliability of the juvenile thoracic cage. From the morphological point of view, a contused lung shows bruising either as a subpleural suffusion or as an intrapulmonary hemorrhage. Lacerations of the lung can result when a severe compressive or crushing force is applied to the chest so that the pulmonary tissue bursts or tears. Another possible mechanism is inward displacement of a fractured rib which impales the lung.

Blunt traumatization of the *heart* manifests itself as concussion, contusion or myocardial rupture. In most cases the force is directly applied to the anterior chest, which compresses or crushes the heart between the sternum and the vertebral column. Bruises of the cardiac wall may be localized in the subepicardial fatty tissue (sometimes in combination with post-traumatic coronary occlusion) or within the myocardium which then appears dark red from interstitial hemorrhage. Lacerations of the heart are most often seen in the relatively thin right ventricle or in the atria; they are less common in the left ventricle, the papillary muscles, the cardiac valves, the interatrial and the interventricular septum. The risk of cardiac rupture is especially high during diastole when the heart chambers are filled with blood and therefore easily burst when they are exposed to a sudden compressive force. Such injuries usually have a fatal outcome either from massive blood loss and hemorrhagic shock (if the pericardial sac is torn and the blood pours into a pleural cavity) or from cardiac tamponade (blood accumulation in the pericardial sac resulting in insufficient filling of the cardiac chambers and impaired forward circulation).

Traumatic aortic ruptures typically occur in vehicular accidents and in falls from a height. The most important mechanism is sudden deceleration, possibly in combination with compression and/or shearing. Traction forces tear the aorta transversely at two sites of predisposition: in the descending part of its arcus (near the attachment of the ligamentum arteriosum) (**Fig. 14**) or immediately above the cusps of the aortic valve. Other locations (for instance in association with a dislocated vertebral fracture) are rather

Figure 14 Transection of the aorta in the distal part of the arch (deceleration injury).

rare. The laceration of the aorta may occur either as a complete or partial transection. In the latter case the outer layers of the vascular wall are not damaged; the intimal tears are often multiple, semicircular and parallel (so-called ladder-rung tears). If the trauma is survived at least for a short time, a parietal thrombosis or a post-traumatic aneurysm may follow as secondary complications.

Abdominal injuries

Blunt force injuries of the abdomen are frequently seen in traffic and work accidents, in child and spouse abuse, in other criminal assaults (with kicking, stamping and punching), but also in suicidal falls from heights. The abdominal organs most vulnerable to blunt trauma are the solid liver and spleen on the one side and the mesentery on the other. Concomitant external signs of blunt traumatization such as contusions or abrasions are by no means obligatory. Substantial injuries to the liver, the spleen and the mesentery always have to be regarded as life-threatening and potentially fatal, especially in cases without rapid surgical treatment. The main reason is internal bleeding into the peritoneal cavity from lacerations (**Figs 15** and **16**). Ruptures of the liver and spleen can be classified as either transcapsular subcapsular laceration. In the first case both capsule and parenchyma are injured so that the blood instantaneously pours into the peritoneal cavity. The second type of laceration is characterized by the initial formation of a subcapsular hematoma which expands continuously and possibly causes a delayed rupture when the covering capsule tears due to overstretching (mostly several hours or even days after the trauma).

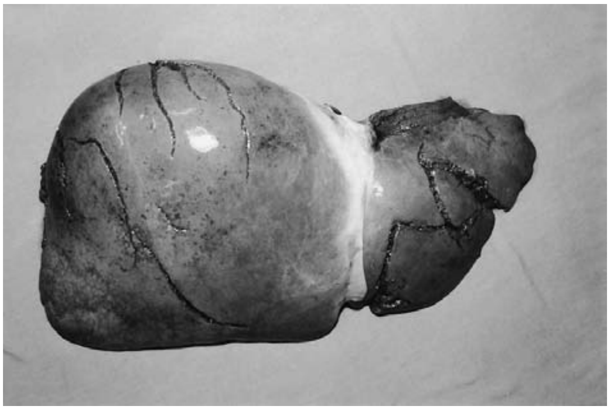


Figure 15 Multiple transcapsular lacerations of the liver.

The stomach and the intestine are less susceptible to blunt traumatization than the parenchymatous abdominal organs. The hollow viscera are more likely to rupture, if they are filled with food or fluid. Another reason why the intestine or stomach might be prone to damage is squeezing of the organs between the indented abdominal wall and the lumbar vertebrae. Fatal outcomes from contusions or lacerations (**Fig. 17**) of the gastrointestinal tract are usually due to diffuse peritonitis.

Renal injuries are a relatively rare source of severe bleeding since the kidneys are deeply located behind the peritoneum. Nevertheless they can be ruptured by a heavy impact to the loin (for instance in traffic accidents or assaults).

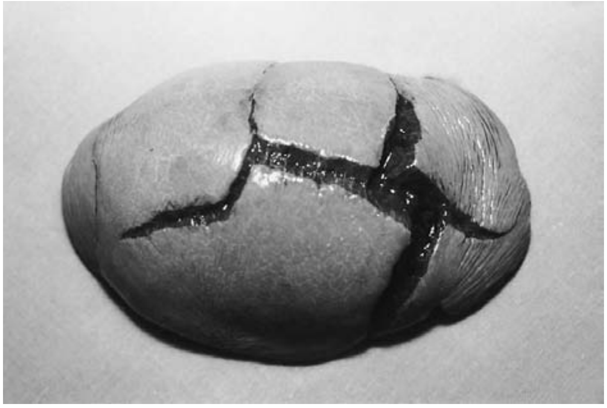


Figure 16 Lacerations of the spleen.

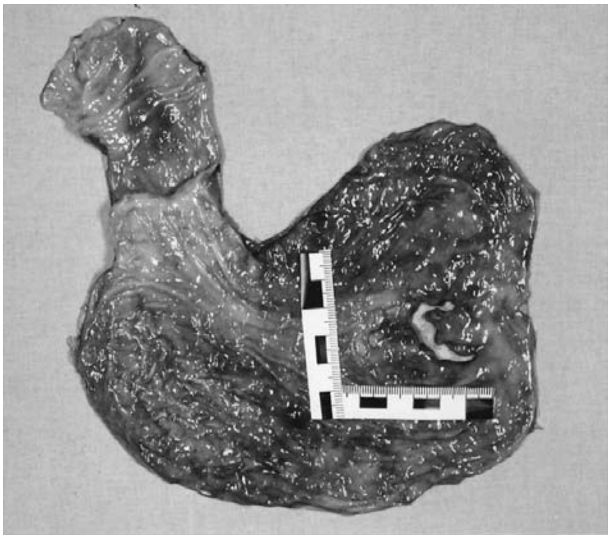


Figure 17 Laceration of the stomach.

Though the empty urinary bladder is placed within the pelvis, when filled it moves upwards and is therefore exposed to blunt traumatization of the lower abdomen. Consequently, rupture of the empty bladder is expected to be extraperitoneal and accompanied with pelvic fractures, whereas a bladder distended with urine may rupture into the peritoneal cavity.

Blunt traumatization of a pregnant uterus is a possible cause of fetal death, mostly due to separation or rupture of the placenta.

Injuries to the Extremities

Apart from injuries to the skin and the subcutaneous layer (see above) other anatomical structures such as the muscles, the bones and the joints may be involved in blunt-force trauma. Extensive crushing of the soft tissues, the formation of blood-filled cavities, comminuted fractures and severence of large vessels are frequent findings in victims of automobile–pedestrian accidents. Internal bleeding (from closed injuries) and external bleeding (from traumatic amputation, severe avulsive wounds and compound fractures) are important factors contributing to hemorrhagic shock. Another sequel to blunt trauma is pulmonary and systemic fat embolism (caused by globules of fat, usually subsequent to fractures or damage of fatty tissues). In cases of prolonged survival, intercurrent infection and pulmonary embolism (originating from post-traumatic venous thrombosis) are dangerous and often fatal complications of an originally non-lethal injury.

See also: **Anthropology**: Skeletal Trauma. **Causes of Death**: Overview; Systemic Response to Trauma. **Clinical Forensic Medicine**: Defense Wounds; Child Abuse; Recognition of Pattern Injuries in Domestic Violence Victims. **Pathology**: Overview; Autopsy.

Further Reading

Ackroyd FW (1977) The wound: assessment by organ systems. The liver, pancreas, spleen and gallbladder. In: Tedeschi CG, Eckert WG and Tedeschi LG (eds) *Forensic Medicine*, vol. I, pp. 210–215. Philadelphia: Saunders.

Buris L (1993) *Forensic Medicine*, pp. 64–95. Budapest: Springer.

DiMaio DJ and DiMaio VJM (1989) *Forensic Pathology*, pp. 87–169. New York: Elsevier.

Hechtman HB (1977) The wound: assessment by organ systems. Gastrointestinal tract. In: Tedeschi CG, Eckert WG and Tedeschi LG (eds) *Forensic Medicine*, vol. I, pp. 215–221. Philadelphia: Saunders.

Ironside JW (1993) Blunt injury to the head. In: Mason JK (ed.) *The Pathology of Trauma*, 2nd edn, pp. 163–177. London: Arnold.

Knight B (1991) *Simpson's Forensic Medicine*, 10th edn. pp. 87–102. London: Arnold.

Knight B (1996) *Forensic Pathology*, 2nd edn. pp. 133–150, 171–230. London: Arnold.

Mason JK (1993) *Forensic Medicine*, pp. 52–61, 72–88. London: Chapmann & Hall Medical.

Mueller B (1975) Schädigungen und Todesfälle durch Einwirkung von stumpfer Gewalt. In: Mueller B (ed.) *Gerichtliche Medizin*, 2nd edn, pp. 383–439. Berlin, Heidelberg, New York: Springer. (in German).

Oehmichen M and König HG (eds) (1997) Neurotraumatalogy: biomechanic aspects, cytologic and molecular mechanisms. *Research in Legal Medicine*, vol. **17**. Lübeck: Schmidt-Römhild.

Prokop O and Radam G (eds) (1987) *Atlas der gerichtlichen Medizin*, 2nd edn, pp. 338–379. Basel: Karger. (in German).

Schulz E (1986) Der Tod durch äuβere Gewalteinwirkung. Verletzung durch stumpfe Gewalt. In: Forster B (ed.) *Praxis der Rechtsmedizin*, pp. 83–106. Stuttgart, New York: Thieme. (in German)

Sherman MM, McCormick JR and Berger LR (1977) The wound: assessment by organ systems. The thorax. In: Tedeschi CG, Eckert WG and Tedeschi LG (eds) *Forensic Medicine*, vol. I, pp. 199–210. Philadelphia: Saunders.

Spitz WU (1993) Blunt force injury. In: Spitz WU (ed.) *Medicolegal Investigation of Death*, 3rd edn, pp. 199–251. Springfield, IL: Charles C Thomas.

Tedeschi CG (1977) The wound: assessment by organ systems. Head and spine. In: Tedeschi CG, Eckert WG and Tedeschi LG (eds) *Forensic Medicine*, vol. I, pp. 29–75. Philadelphia: Saunders.

Unterharnscheidt F (1993) *Pathologie des Nervensystems. VI. A. Traumatische Schäden des Gehirns (forensische Pathologie)*. Berlin, Heidelberg: Springer. (in German)

Burns and Scalds

D J Pounder, Department of Forensic Medicine, University of Dundee, Dundee, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0431

Introduction

A burn is an injury caused by heat or by a chemical or physical agent having an effect similar to heat. Most burns are produced by dry heat, and result from contact with a flame or a heated solid object, or alternatively from exposure to the radiant heat of an object. Burns caused by moist heat are described as scalds. Chemical burns are produced by acids and alkalis, or by vesicants especially developed as chemical warfare agents. Microwaves and electricity also produce burns.

Cause and Manner of Death

The severity of a burn caused by dry heat is assessed by two parameters, firstly the depth of the burn injury, and secondly the extent of injury, that is the size of the burn relative to the total body surface area. In assessing burn severity, it is helpful to keep in mind that a common error is to underestimate the depth and to overestimate the extent. In addition to burn depth and extent, other factors determining mortality are the location of the burn, the age of the victim and the presence of other injuries or natural disease. In general, burns involving more than 50% of the body surface carry a poor prognosis. However, age is a major factor for children under 2 years and adults over 60 years, so that in these victims more than 20% surface area involvement carries the same poor prognosis. In the first 48 h period after burning the major threats to life are hypovolemic shock and shock-induced organ failure, primarily renal failure.

The majority of burn deaths are related to accidental domestic fires, in which children and the elderly are particularly at risk, as a result of carelessness coupled with an inability to effectively combat or escape the fire. Alcoholics and drug addicts are a third at-risk group. Suicide by burning is rare, likely because of the pain involved. It is more common in eastern cultures than in the west, and the ritual suicide of a Hindu widow on the funeral pyre of her husband, Suttee, is now prohibited in India. During the period of the Vietnam war, the self-immolation of Buddhist monks during political protests was shown on television worldwide. Rarely a homicide victim may be doused in flammable liquid and then set alight. Petrol bombs, or Molotov cocktails, are improvised incendiary devices which have become popular in urban disturbances, but do not usually result in death. Necklacing is a method of homicidal burning developed in South African black townships during the apartheid era as a form of punishment for political opponents. It involves placing a vehicle tire around the neck of the victim (hence 'necklacing') and setting it alight.

Burns resulting from flaming clothing have a pattern reflecting both the nature of the clothing and the position and movements of the victim. The type of fabric can influence the burn severity. The area and depth of burn tends to be greater with the faster-burning cotton fabrics than with polyester. Higher fabric weight reduces burn severity, and woven fabrics are associated with larger burns than knitted fabrics. Flash burns result from sudden deflagrations burning exposed skin surfaces which are not protected by clothing. These uniform, typically partial thickness, flash burns may be then partly masked by subsequent flame burns from ignited clothing. Burns produced as a result of contact with a hot solid object often leave a brand-mark in the shape of the object, for example the triangular base of an iron. Burns from cigarettes are of the expected size, round and punched out. To produce cigarette burns requires firm contact for some seconds, and cannot occur simply as the result of the accidental dropping of a cigarette, or brushing against one. A cigarette burn implies deliberate infliction, more obviously so when multiple burns are present. They may be seen in victims of child abuse, torture in custody, interprisoner violence and as the result of self-harm in individuals with low self-esteem and personality disorders. Whether fresh injuries or old scars, cigarette burns seen at autopsy always raise serious concerns which demand further investigation.

Burn Severity

The traditional classification of burn depth is into three degrees. A first degree burn destroys only the epidermis. It is characterized by erythema, edema and pain. Sunburn, produced by the radiant heat of the sun, is the most common first degree burn. In general, first degree burns are produced by prolonged exposure to low intensity heat or very brief exposure to high intensity heat. Healing, associated with skin peeling, is usually uneventful and completed in 5–10 days with no residual scarring.

A second degree burn involves both the epidermis and a variable depth of the underlying dermis. The most superficial second degree burns implicate only the upper third of the dermis and are characterized by

blister formation. They are extremely painful but heal in 7–14 days with no or minimal scarring. A deep second degree burn extends beyond the upper third of the dermis but not beyond the dermis itself. Paradoxically these deeper burns are less painful as a result of destruction of nerve endings in the dermis. Healing is extremely slow, sometimes requiring months and usually leading to dense scarring, if not treated by skin grafting.

A third degree burn destroys the full thickness of the epidermis and dermis. Heat coagulation of dermal blood vessels leaves the tissue avascular with a characteristic waxy white appearance. If there is prolonged contact between subcutaneous fat and flame then the burn has a leathery brown or black, charred appearance. There is characteristic lack of pain, due to heat destruction of all nerve endings. Spontaneous regeneration of skin (primary re-epithelialization) will not occur and such burns require skin grafting. Burns can also be classified, according to the modern classification, as partial thickness or full thickness, depending on the depth of skin involved. Beneath any burned tissue there is usually a zone of marginally viable tissue which is readily converted into nonviable tissue by physiological stressors. In this way a second degree burn frequently converts into a third degree burn.

The second parameter of burn severity is the extent of injury. This is expressed as the percentage of the total body surface area which is burnt. It can be estimated using the 'rule of nines' which divides the body into areas representing 9% or multiples of 9% of the total body surface. Thus the head and neck are 9%, each arm 9%, each leg 18%, the anterior trunk 18%, the posterior trunk 18%, and the genitalia and perineum 1%. In making the assessment a very rough guide is that the victim's palm is approximately 1% of the total body surface area. In children under 15 years body proportions differ and the estimates must be age-adjusted.

Bodies from Fires

When a body is recovered from a fire, a common problem is to distinguish burns produced during life from postmortem burning of the corpse. Most fire deaths result from carbon monoxide poisoning and smoke inhalation, and the burns to the body are postmortem, but this generality does not assist in resolving the issue in an individual case. It is commonly impossible to exclude burning during life, firstly because postmortem burns may be superimposed on and mask antemortem burns, and secondly because antemortem burns may be indistinguishable from postmortem burns. When the interval between burning and death is short there may be insufficient time for the development of the typical red margin of vital reaction around the burnt area. If time is sufficient then the margin is pronounced and distinctive, measuring 1–2 cm in width. Postmortem burns may be associated with marginal congestion, thought to be the result of tissue contraction, but this is typically a narrow band only a few millimeters wide. In the face of these diagnostic difficulties, and the social pressure to minimize the perceived suffering of the deceased in the eyes of the next of kin, most of these burns are dismissed, rightly or wrongly, as postmortem.

The burning of a corpse produces other artifacts which may create confusion. Characteristically, heat contraction of the muscles causes the limbs to flex and the body to assume a pugilistic (boxer's) posture, an appearance which does not reflect in any way a struggle at the time of death. Desiccation and loss of the distal limbs may give a false impression of small body size. Heat contraction of the skin produces splits, which are distinguished from wounds produced in life by the fact that they involve charred areas of skin, are uniformly shallow and show no evidence of hemorrhage, as a consequence of heat coagulation of the blood vessels. Contraction of the skin of the torso may induce a prolapse of the rectum and vagina, and contraction of the neck and face forces protrusion of the tongue. Bones may be fractured by the intensity of the heat, but the fractures are unlike those produced by trauma during life. Complex right-angled fractures of long bones ('street and avenue fractures') and repeated arcs of fractures are characteristic. Burning of the scalp produces not only an eggshell fracturing of the skull vault but also a postmortem extradural hematoma. This is thought to result from blood boiling in the diploe of the skull and then being forced into the extradural space.

Scalds

Scalds are produced by moist heat which may be steam or any hot liquid, such as water, oil or even molten metal. They are typically less severe than burns produced by dry heat. The scalded area appears erythematous with desquamation and blistering of the usually sharply demarcated area of injury. Unless the liquid is superheated, such as oil, there is no singeing of hairs or charring and carbonization of tissues. Scalds generally occur on exposed skin. In clothed areas the influence of the clothing on the severity of the injury is variable. Fabrics which are poorly permeable protect the skin, whereas absorbent fabrics may hold the hot liquid against the skin, reduce natural cooling, and make the scald more severe.

The pattern of the scald can give an indication as to how it occurred. Major areas of scalding are sharply demarcated with trickle marks reflecting the flow of hot liquid under the influence of gravity. There may be splash marks. Dipping injuries of the limbs appear as well-demarcated glove and stocking scalds. Distinguishing accidental from deliberately inflicted scalds, by evaluating the pattern of injury against the alleged circumstances, is of particular importance in childhood injuries, both fatal and nonfatal. Industrial accidents involving superheated steam produce severe scalds in a pattern similar to flash burns.

Chemical Burns

Chemical burns are produced by corrosive acids and alkalis. The pattern on the skin surface resembles that of scalds caused by liquids, but the injuries differ in physical appearance, reflecting different mechanisms of tissue damage. The amount of tissue damage caused in a chemical burn will depend on the strength of the chemical, its concentration, the quantity applied to the skin surface, the duration of contact, and the extent to which the chemical penetrates the tissues. Chemical burns continue to develop for as long as the causative agent is not neutralized by another chemical, or inactivated as a result of reaction with the body tissues.

Acids with a pH less than 2 precipitate proteins causing coagulation necrosis. The resultant burns are clearly demarcated, dry and with a leathery scab, the color of which depends on the acid. Nitric acid gives a yellow–brown scab, sulfuric acid (vitriol) a black–brown scab, hydrochloric acid (spirit of salt) a white to gray scab, and carbolic acid (phenol) a light-gray to light-brown scab. Prolonged contact with cement may produce a chemical burn since the pH of cement is less than 2. Alkalis with a pH above 11.5 also produced chemical burns. They generally produce more tissue damage than acids because they cause liquefactive necrosis, which facilitates ever deeper penetration of the alkali. The caustic alkalis and ammonium hydroxide leave a gray–white mucoid burn.

Chemical burns from acids and alkalis are almost invariably accidental, and homicide by this method is rare. Throwing liquid corrosives, such as acid, over a victim is more often intended to produce facial disfigurement than death, and to be the act of a spurned suitor. Suicide by the ingestion of strong acid or alkali has now become rare in the developed world because of the ready availability of drugs, which are painless. However, suicide by this method is still seen in poorer countries. Typically there is staining of the lips, and often the cheeks, chin and neck, as well as chemical burns of the mucosa from lips to stomach, sometimes extending into the small bowel. Esophageal and gastric perforations are most common with sulfuric and hydrochloric acids.

Chemical Warfare Agents

One group of chemical warfare agents produce skin burns. The best known of these vesicants is 'mustard gas', or more correctly sulfur mustard. It was given its name because of the smell, which has been likened to garlic, mustard, horseradish or leeks. Mustard gas was first used by Germany during World War I, at Ypres. Since then it has been used by the Italians in Ethiopia in 1936, the Japanese against Chinese troops during World War II, and Iraq against Iranian troops in the 1980s. During World War I total mustard gas casualties may have been as high as 400 000, but with a low mortality rate of around 2–3%. Death within 24 h of exposure is very rare, whereas later deaths are the result of respiratory effects or bone marrow depression with associated infections. It was the incapacitating, rather than the lethal, effects of mustard gas that caused it to be named the 'King of the Battle Gases' during World War I.

The vapour given off by liquid sulfur mustard rapidly penetrates clothing to damage the underlying skin. The production of blisters is a characteristic of mustard gas, by contrast with acids and alkalis which do not produce blistering when they burn the skin. The mechanism of blister formation is unknown, but it is a specific response of human skin which is not seen in animal studies. Following exposure to sulfur mustard there is a latent, symptom- and sign-free period of some hours. If the exposure is severe, the first sign is erythema, reminiscent of scarlet fever. There is mild skin edema, and itching is common. As the erythema fades, areas of hyperpigmentation are left behind, as occurs with sunburn. By 18 hours typical vesicles appear. These blisters are uncomfortable but not painful in themselves, although they are delicate and easily rubbed off, leaving the now painful blister base. Crops of new blisters may appear as late as the second week after exposure. Analysis of the aspirated blister fluid for thiodiglycol may assist in distinguishing mustard gas blisters from those produced by other chemical warfare vesicants, such as lewisite. Deep burning, leading to full thickness skin loss, is followed by eschar formation. The effects of liquid sulfur mustard on the eye mirror those in the skin. Late onset blindness is a distressing effect of exposure.

Lewisite is another chemical warfare agent which, like mustard, is a vesicant when applied to human skin but not to animals. Unlike mustard gas it is

nonpersistent, making it possible for attacks to be launched on previously contaminated ground. However, lewisite is rapidly hydrolyzed when mixed with water, making it ineffective as a weapon under wet conditions. Although the chemical warfare effectiveness of both lewisite and mustard depend on their vesicant properies, lewisite is also lethal. Less than 3 g applied to the skin of an average man, and not washed off or otherwise decontaminated, would be expected to be fatal. Death is typically the result of damage to the respiratory tract, with necrosis of the mucosa and formation of a false diphtheria-type membrane, together with pulmonary edema and congestion, with secondary bronchopneumonia. It has not been used in warfare and deaths have resulted from accidental exposure. A related arsenical vesicant, phenyldichlorarsine, was used on a large scale during World War I because it was capable of penetrating the respirators then available. Other vesicants similar to lewisite are ethyl-phenyldichlorarsine (known as dick) and methyl-phenyldichlorarsine (known as methyl-dick).

Electrical Burns

The principal bodily barrier to an electrical current is the skin, and once beyond the dermis the current passes easily through the electrolyte-rich fluids. Having entered the body at the entry point, usually the hand, the electricity then exits to earth (ground) via a pathway depending mainly on the relative resistance of the various potential exit points. The current tends to take the shortest route between entry and best exit irrespective of the varying conductivity of the different internal tissues. Alternating current (AC) is more dangerous than direct current (DC), and AC in the range 39–150 cycles per second has the highest lethality. The effects of AC depend on the magnitude, frequency and duration of the current, whereas the voltage is of importance only because it is a factor in determining the current. DC injuries are uncommon but examples include encounters with lightning, car batteries, electroplating, some public transportation systems and some industrial systems. The mechanism of death in electrocutions is most commonly a cardiac dysrhythmia, usually ventricular fibrillation, less commonly paralysis of the respiratory muscles, and rarely a direct effect on the brainstem as a result of passage of the current through the head and neck. Hand-to-hand passage of a high-voltage current has a reported immediate mortality of 60% as a result of cardiac arrhythmia.

Skin burns are a common form of electrical injury and are a pathognomonic marker for death by electrocution. When present, the typical skin lesion is a thermal burn resulting from the heating of the tissues by the passage of the electric current. Tissue damage from this heating effect may be insufficient to produce a visible injury if the surface contact area is broad and the conductivity of the skin is high because of a high water content, two conditions which are usual in electrocutions in the bath. Torture by electricity may be performed using broad wet contact electrodes in order to avoid leaving evidential marks. When they occur, electrical burns on the skin may be of firm contact or spark (arc) type. Both types can occur in the same victim as a result of the irregular shape or movement of the conductor, or movement of the victim during electrocution.

A firm contact electrical burn of entry typically leaves a central collapsed blister, which may reproduce the shape of the conductor, with a surrounding areola of pallor. The blister is created by the steam produced in the heating of the tissues by the electric current. When the current ceases, the blister cools and collapses to leave a crater with a raised rim. Should the blister burst during its formation, as a result of its large size or the continued passage of current, then the epidermis may peel off leaving a red base. Contact electrical burns at exit points are often not seen but should be searched for. When present, in low voltage deaths, they are similar to, but less severe than, the corresponding entry mark. In high-voltage (more than 1000 volts) electrical burns the contact injury of exit often appears as a 'blow-out' type of wound. Skin and subcutaneous tissue may be destroyed, exposing thrombosed vessels, nerves, fascia, bones or joints.

With the passage of an electric current, metallic ions from the metal conductor combine with tissue anions to form metallic salts which are deposited in tissues, and may be demonstrated by chemical, histochemical and spectrographic techniques. The histological appearance of electrical skin marks is closely similar to thermal injuries with cellular eosinophilia and nuclear streaming. Some researchers have claimed to be able to distinguish, at the histological level, electrical from thermal injury, but this is disputed. Certainly the combination of the gross appearance and histology will usually permit a firm diagnosis.

A spark (arc) burn occurs when there is an air gap between the conductor and the skin so that the current arcs across the gap as a spark. The distance which the spark can jump is proportional to the voltage, so that 1000 volts can jump a few millimeters, 5000 volts can jump 1 cm, and 100 000 volts can jump 35 cm. The extremely high temperature of sparks, which may be up to 4000°C, causes the epidermal keratin to melt over a small area. After cooling, this leaves a raised brown or yellow nodule of fused keratin surrounded by an areola of pale skin. A brief arc transmits only enough energy to cause a

superficial skin burn. These are most commonly seen on the hands. Ocular burns, mostly caused by low voltage arcs, are a particular clinical problem in survivors. High voltage spark burns may cause large areas of skin damage resulting in an appearance of 'crocodile skin'. Spark burns involving the clothing can cause the clothing to ignite, so that the victim suffers superimposed flame burns.

The severity of the electrical injury to the deep tissues depends on the amperage, i.e. the actual amount of current passing through the tissues. Although it is impossible to know the amperage, it can be inferred from the voltage of the source as either high or low. A low-voltage household source is capable of causing death if a sufficient current passes through the body and 60 mA will produce cardiac fibrillation. However, no deep tissue damage is evident at autopsy because the current pathway is too diffuse to cause thermal damage. Consequently, there are no characteristic internal findings in fatal electrocutions. Skeletal fractures and joint dislocations may occur as a result of tetanic contractions. Skeletal muscle damage leads to release of myoglobin, and muscle-specific intracellular enzymes, with resultant myoglobinemia and myoglobinuria. A high-tension source producing a current of 5000 mA or more is usually required to produce severe widespread tissue necrosis. Experimental studies have shown that this tissue necrosis is a result of not just heat but also short-term nonthermal effects of electric fields. Although the severity of injury is directly proportional to the duration of the current flow, even very brief exposures to high amperage will produce massive deep tissue damage. These types of electrical injuries are more akin to crush injuries than to thermal burns in as much as the damage below the skin is usually far greater than the outward appearance would indicate. If, after death, the electric current continues to flow then there may be severe damage to the body with peeling and blistering of the skin, charring, and cooking of the underlying tissues. Rarely, thermal burns of this type may be found in a survivor, usually following prolonged contact with over 1000 volts. In these cases, necrosis of deep tissues, both immediate and delayed, often requires limb amputation. Generally, for those who survive an electric shock, the prognosis is good and the majority make a complete recovery, so that delayed deaths from electrocution are uncommon.

Most electrocutions are accidental and the bathroom is a place of particular danger in the home. An unusual and characteristic finding in electrocution in the bath is that the subsequent development of hypostasis is limited by the water-line, resulting in a stark and unusual demarcation. Currents of less than 0.2 mA will not cause a skin injury or death by electrocution but are sufficient to evoke a startle reaction and may precipitate a lethal accident, such as a fall from a height. Suicidal electrocution is uncommon but increasing, and may be difficult to distinguish from an accident. Homicidal electrocutions are also rare, except in a judicial setting, the first execution by electricity having been carried out in Auburn Prison, New York in 1890.

Lightning

Each year, lightning causes hundreds of deaths worldwide. A lightning bolt is produced when the charged undersurface of a thunder-cloud discharges its electricity to the ground. Very large electrical forces are involved with currents up to 270 kA. The lightning may directly strike the victim, strike a nearby object and then jump from the object to the victim (a side flash), or strike an object which conducts the lightning to a victim in contact with the object, e.g. a worker in contact with a metal crane. The lightning current may spread over the body surface, pass through the body or follow both routes. The resulting injuries may be electrical, burns, or blast from the wave of heated air created by the lightning strike. The pathological findings range from a completely unmarked body to bizarre and extreme trauma. An unmarked dead body found in the open should raise the possibility of a lightning strike. When injuries are present, the clothing may be torn and the shoes burst, with the body partly stripped and the clothing scattered. The hair may be seared, patterned skin burns reflect the heating of metal objects such as zippers, severe burns with blisters and charring may be present, and rupture of the tympanic membranes, fractures and lacerations may be found. Fern-like or arborescent patterns on the skin are pathognomonic of a lightning strike but are rarely seen. They appear as irregular red marks, several inches long, which tend to follow skin creases and the long axis of the body.

See also: **Explosives**: Analysis. **Fire Investigation**: Types of Fire; Physics/Thermodynamics; Chemistry of Fire; Evidence Recovery at the Fire-scene; Fire-scene; Fire-scene Patterns; Laboratory.

Further Reading

Demling RH and LaLonde C (1989) *Burn Trauma*. New York: Thieme Medical.

Lee RC, Cravalho EG and Burke JF (1992) *Electrical Trauma: the Pathophysiology, Manifestations and Clinical Management*. Cambridge: Cambridge University Press.

Marrs TC, Maynard RL and Sidell FR (1996) *Chemical Warfare Agents: Toxicology and Treatment*. Chichester: John Wiley.

Postmortem Changes

M J Lynch, Victorian Institute of Forensic Medicine, Monash University, Melbourne, Victoria, Australia

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0425

Introduction

The changes that occur in the human body after death may be classified and contrasted in a number of ways. There are the predictable, genuine and natural versus the artifactual. The latter may be deliberately produced or occur accidentally. There are changes which assist the investigator and those which can confound. And the changes occur at all levels of pathological investigation; macroscopically, microscopically, biochemically and toxicologically. The challenge for the forensic pathologist lies in drawing the appropriate inferences from the genuine changes, recognizing the artifacts and interpreting within the limits of the postmortem examination the available information.

The Purpose of the Postmortem Examination

The precise requirements of the medicolegal autopsy will vary from jurisdiction to jurisdiction. However, there are certain fundamental issues which the forensic pathologist is usually required to address. These include identification of the deceased, determination of a cause and manner of death, estimation of a time of death and interpretation of any injuries which might be (or have been) present. The various categories of postmortem change (**Table 1**) are examined here in some detail.

Table 1 Categories of postmortem change

Genuine	Decomposition	Putrefaction Mummification Adipocere formation Autolysis
	Rigor mortis Hypostasis	
Artifactual	Predators Embalming Undertaker's fracture Muscle rupture Resuscitation artifact 'Bruising'	
Time of death	Temperature Rigor mortis Hypostasis Gastric emptying Entomology	

Genuine Postmortem Changes

Decomposition

Decomposition of a human body is a natural and predictable sequela of death. The postmortem examination of such a body is among the more challenging tasks confronting a forensic pathologist. It is at such times that the pathologist needs to be on guard lest some crucial piece of evidence be overlooked, misinterpreted or even inadvertently destroyed.

Decomposition can be conveniently subclassified into three basic processes: putrefaction, mummification and adipocere formation. The different processes often co-exist and the rate and extent of their development is influenced by the prevailing environmental conditions such as temperature, moisture level, presence or absence of concomitant insect activity and whether the body is submerged or buried.

Postmortem hypostasis

This phenomenon represents the gravitational shifting of blood which will occur as a natural and predictable postmortem phenomenon. Its forensic importance is twofold. First, the investigator must be aware that postmortem hypostasis or lividity will become fixed after a variable time. A figure often quoted is approximately 12 h. However, the use of such a time frame needs to be exercised with caution. In the author's experience it is the appearance of hypostasis in autopsy photographs from suspicious deaths which invariably attracts the attention of legal counsel in court proceedings, who often confuse the changes with some form of injury given the potential similarity to bruising. The one area where postmortem hypostasis may provide valuable information to the forensic investigator is in determining whether a deceased person has been moved after death. If the pattern of lividity is inconsistent with gravitational shifting of blood, one can infer that the deceased has been moved postmortem and that the shifting of the body has occurred after lividity became fixed.

Rigor mortis

The development of postmortem rigor mortis is a predictable postmortem response. Biochemically it represents postmortem muscle contraction which becomes fixed due to the inevitable diminution of available adenosine triphosphate (ATP). It appears first in the smaller central muscle groups such as the jaw and

face and moves peripherally. It declines in the reverse order. Some textbooks provide a framework within which to estimate time of death based on development of rigor mortis. However, these are crude estimates and furthermore their use in assisting with determination of time of death is confounded by the notion of 'cadaveric spasm'. This is a phenomenon whereby in an individual engaged in some form of intense physical activity immediately prior to death, the development of rigor can occur as an immediate postmortem response.

The changes which occur in the skeletal muscle postmortem are chemical reactions and as such may be accelerated by an increase in temperature. Thus a body that has been exposed postmortem to high environmental or ambient temperatures may develop rigor at a more rapid rate than expected. Similarly, the death of an individual occurring in a cold environment or refrigeration immediately postmortem will delay the onset of rigor mortis.

Autolysis

This represents the natural breakdown process of human tissue which will be facilitated by bacterial activity. A number of specific postmortem artifacts are worthy of mention here as they may cause confusion with specific antemortem pathological processes. So-called postmortem esophago- or gastromalacia may simulate rupture of these visci. Breakdown of the pancreas post mortem can cause confusion macroscopically with acute hemorrhagic pancreatitis. Exaggerated hypostasis involving the posterior left ventricle in an individual who has remained supine after death can cause confusion with ischemic myocardial damage. And finally in bodies recovered from fire the well-recognized phenomenon of heat associated extradural hemorrhage, producing the characteristic honeycomb appearance of blood which has exuded from the diploic veins and venous sinuses, is worthy of emphasis.

Putrefaction

In warm climates this change is the norm and can occur quite rapidly over even a matter of hours. The order in which changes appear is relatively predictable. The first sign is usually a green discoloration of the skin of the anterior abdominal wall in the right iliac fossa region. This is a reflection of the relatively superficial location of the large bowel (cecum) and its invariably high bacterial content. The presence of so-called marbling then follows. This represents bacterial overgrowth in superficial veins and confers a 'road map' type pattern on the body, often first appearing in the shoulder regions.

A series of predictable changes then ensues. With increasing postmortem gas formation the body swells, in particular the abdominal cavity and scrotum. There is a purging of bloodstained fluid from the nostrils and mouth often raising concerns among uninitiated investigators of suspicious circumstances surrounding the death. If there has been concomitant insect activity with flies laying eggs there will often be intense entomological activity with numerous maggots present on and about the body.

Adipocere formation

This particular type of postmortem change comprises the appearance of a white waxy material on the surface of the body. Its name is derived from adipo- (fat) and -cire (wax). This commonly occurs in bodies that have been exposed to moisture or submerged and where the temperatures are cool. Biochemically it represents breakdown products of fatty acids and has been shown to contain palmitic, oleic and stearic acid. Although this process is less malodorous than putrefactive change it can have the effect of concealing identifying features and injuries and, depending on the jurisdiction in which one works, might be a change rarely encountered.

Mummification

Mummification is the type of change most familiar to the lay person. It invariably occurs when the body has been in a dry environment without exposure to insect activity. The drying of the skin can markedly hamper identification and complicate injury interpretation although often injuries will be preserved relatively intact.

Postmortem Changes Impinging on Time of Death

These represent changes which are predictable and potentially helpful to the pathologist. The limitations however on the use of the various methods of estimating the postmortem interval needs to be remembered.

Rigor mortis and postmortem hypostasis

The potential benefit of these two processes and their limitations with respect to the determination of postmortem interval have been noted above.

Body temperature

The core temperature of the human body post mortem will fall in a semipredictable way. The investigating pathologist needs to be aware of the potential usefulness of a postmortem temperature but more importantly of the limitations on interpretation of such data.

The most widely accepted method is using Henssge's nomogram which requires the taking of a rectal temperature. Alternative sites for temperature measurement include outer ear and abdominal viscera.

Postmortem chemistry

Forensic literature abounds with material on the subject of postmortem biochemical analysis. The most useful specimen for such postmortem analyses is vitreous humor which is relatively isolated from post mortem bacterial overgrowth.

A number of general comments may be made about the usefulness of certain postmortem vitreous measurements. Vitreous urea and creatinine are stable postmortem and thus provide good markers for the assessment of antemortem dehydration (elevated urea coupled with normal creatinine) or renal impairment (variable elevation in both urea and creatinine). The adjunctive use of postmortem electrolytes is more problematic. Vitreous sodium and chloride will generally fall but at a variable rate whereas vitreous potassium will rise. It is the latter quasi-predictable event which has attracted the particular attention of academics.

With respect to estimation of postmortem interval the measurement of vitreous potassium is of most value after one to two days. Formulae have been proposed to assist in the interpretation of postmortem interval. That of Madea and Henssge is:

$$\text{Postmortem interval (PMI(h))} = 5.26 \times \text{potassium concentration (mmol/l)} - 30.9$$

and that of Sturner is:

$$\text{Postmortem interval (PMI(h))} = 7.14 \times \text{potassium concentration (mmol/l)} - 39.1.$$

It needs to be reiterated that caution is required in the use of such equations and in the interpretation of results.

Another method by which postmortem interval may be estimated is the measurement of vitreous hypoxanthine. This is attracting increasing attention in the forensic literature. A comparison of the use of postmortem vitreous hypoxanthine and potassium levels concluded that a combination of the two generated the greatest accuracy with respect to estimation of postmortem interval.

Gastric emptying

This subject is included more for the sake of completeness than for any useful contribution to the determination of postmortem interval. A perusal of review articles which examine the factors which delay, accelerate or have no effect on gastric emptying and an awareness of gross miscarriages of justice makes one uneasy especially when even today experienced pathologists are seen to prognosticate about a particular time of death in criminal proceedings based on gastric emptying. The only useful observation that can be made with respect to gastric contents is the nature and volume of the contents which may provide an important piece of information with respect to a criminal investigation. The number of variables otherwise associated with the process of gastric emptying make this an area that should be approached with a combination of extreme caution and healthy skepticism.

In conclusion, various methods exist to aid the pathologist with respect to the determination of the time of death. Some of these are truly academically based and only used with any frequency in specific jurisdictions. The interested reader seeking information concerning postmortem skeletal muscle electrical stimulation and additional detail on the foregoing is referred to the literature.

Artifactual Postmortem Changes

Postmortem injuries by predation

The predators in question can range from the domesticated to the wild and from the smallest members of the insect kingdom to the largest of carnivorous mammals. Postmortem insect activity in the form of fly larvae or beetles is a major contributor to the natural breakdown process of the human body after death. Intense larval activity as well as obscuring distinguishing marks and features, stigmata of natural disease and signs of injury also produces markings on the skin in the form of circular or elliptical holes which can 'mimic' injury to even the trained observer. Such postmortem insect activity although generally tending to obscure can be of assistance to the investigating pathologist with respect to determination of the time of death. The pathologist needs to ensure that appropriate specimens are obtained and the engagement of a forensic entomologist, with knowledge of the specific insect life cycles, might also provide valuable information.

In circumstances where an individual has died in a domestic dwelling and remained undiscovered for some days, injuries present on the body may ultimately be attributable to the family pet. Injuries produced by domestic dogs are not uncommonly seen in forensic practice. In individuals dying in more dilapidated surroundings the contribution of incised injuries by small mammals such as rats may also warrant consideration. Information concerning the scene of discovery is of crucial importance in such circumstances.

The final contribution of predators may occur when the individual is exposed to large carnivores.

The author has the experience of a case where an individual entered a lion enclosure at a metropolitan zoo (without authorization) in order to engage in some form of martial art ritual with the captive beasts. A large numbers of injuries were observed on the deceased which were considered to represent postmortem artifact. The unfed animals had essentially stripped skeletal muscle from the long bones while leaving much of the abdominal and thoracic viscera intact. The circumstances of the case made the interpretation of the injuries relatively straightforward however, one needs to be prepared for a situation where relevant information is not so forthcoming.

'Perimortem' injuries

Injuries inflicted deliberately postmortem, invariably in the context of homicide, are encountered not infrequently by the practicing forensic pathologist. These generally fall into two broad categories. Those produced in a so-called 'perimortem period' and those produced clearly postmortem.

With respect to perimortem injuries often the pathologist in court when describing their findings with respect to multiple homicidal stab wounds is forced to concede that, due to the absence of bleeding associated with some of the injuries, there is the possibility that these were produced after death, i.e. after the 'irreversible cessation of blood flow or irreversible cessation of brain function' (statutory definition of death in the author's jurisdiction). In such instances the degree of bleeding adjacent to the specific wounds is of some assistance but such issues generally cannot be addressed with absolute certainty.

Postmortem mutilation

The second category of injuries warranting consideration is those produced in a ritualistic sense generally signifying significant psychopathology in the assailant. The type and degree of such postmortem mutilation varies greatly but includes removal of breasts, genital mutilation such as removal of the penis and scarification type injuries. The latter may sometimes cause confusion with antemortem injuries produced in the context of torture.

Injuries related to handling of the body

Injuries in this category include the so-called 'undertaker's fracture' whereby careless or overly rough handling of the deceased person results in a fracture of the cervical spine. Correct interpretation of this is generally facilitated by the absence of bleeding although sometimes a small amount of hemorrhage may be present which causes difficulties. In such instances, as in any forensic autopsy, the forensic pathologist needs to rely on the circumstances surrounding the death including the treatment of the deceased person after discovery. A related artifact is the rupture of skeletal muscle which can occur by the forceful, albeit naive, breaking of rigor mortis, usually by inexperienced funeral directors or forensic technicians. In such instances examination of the injured skeletal muscle usually reveals a small amount of hemorrhage but obviously no evidence of an inflammatory reaction.

Prinsloo Gordon artifact

This eponymous entity is well known to most experienced forensic pathologists. It represents hemorrhage on the anterior aspect of the cervical spine, posterior to the trachea and esophagus. This may be considered a form of hypostasis. It causes most confusion in the context of the interpretation of neck injuries. It is minimized by insuring that the standard practice in cases of neck injury involves either opening the cranial cavity first to allow drainage of blood or alternatively if the extracranial examination is to occur first, insuring that the root of the neck is incised to allow drainage of blood from that region.

Postmortem 'bruising'

From a semantic viewpoint, if one considers a bruise to be an active physiological response to tissue trauma then this category is probably better described as postmortem extravasation of blood potentially simulating antemortem bruising. A number of difficulties can arise with respect to the interpretation of such changes. Histopathology or enzyme histochemistry may sometimes be of assistance. Most important is an awareness of this type of artifact. The potential problems associated with such cases are discussed in a problematic case reported by Burke et al.

Resuscitation injuries

Postmortem artifacts from a number of sources can provide confounding data for the forensic pathologist. Ideally, before commencement of the postmortem examination the circumstances surrounding the death will have been made clear to the pathologist and this will include specific descriptions of medical intervention and resuscitation, both professional and inexpert. This, however, is invariably the exception rather than the rule. Resuscitation artifact is particularly problematic in pediatric homicide cases where rupture of abdominal viscera, mesenteric laceration and retinal hemorrhages may be the subject of intense and rigorous cross-examination in the later criminal trial. Rib fractures, laryngeal fractures, cardiac rupture and pneumoperitoneum are all well-described entities.

Embalming artifact

This category of postmortem change is worthy of mention for two reasons. First, the sites of introduction of embalming fluid can mimic penetrating injuries of more sinister origin. Although the fact that an individual has been embalmed would generally be known to the investigating pathologist, it is often in cases of second postmortem examination particularly in a deceased individual who has died in a foreign jurisdiction that the pathologist needs to be aware of the potential for such confounding artifacts. Secondly, postmortem toxicology on embalmed individuals can produce quite puzzling results with high levels of methanol, anticoagulants and various other dyes often detected by sophisticated screening methods.

Conclusion

The significance of the postmortem changes described above lies in their ability to confuse and obscure. A practicing forensic pathologist needs to be well focused on the purpose of the medicolegal autopsy. He or she needs to be aware of the different categories of postmortem change and to recognize when they can be of assistance in the investigation, when they might assist but need to interpreted with caution and when they will do little but obscure. The forensic pathologist is in a privileged position each time they embark on a postmortem examination. It is their responsibility to the profession, to the medicolegal investigating authority and to the family of the deceased that they elicit the maximum information from the autopsy and interpret it to the best of their ability.

See also: **Toxicology**: Methods of Analysis – Post Mortem. **Pathology**: Postmortem Changes; Postmortem Interval.

Further Reading

Burke MP, Olumbe AK, Opeskin K (1998) Postmortem extravasation of blood potentially simulating antemortem bruising. *American Journal of Forensic Medicine and Pathology* **19**:46–49.

Coe JI (1993) Postmortem chemistry update, emphasis on forensic application. *American Journal of Forensic Medicine and Pathology* **14**:91–117.

Fatteh A (1972) Artefacts in forensic pathology. In: Wecht C (ed.) *Legal Medicine Annual*, pp 49–74.

Henssge C, Knight B, Krompecher T, Maedea B and Nokes L (1995) *The Estimation of the Time Since Death in the Early Postmortem Period*. London: E Arnold.

Human Tissue Act, 1982 (Vic) Section 41.

James RA, Hoadley PA and Sampson BG (1997) Determination of postmortem interval by sampling vitreous humour. *American Journal of Forensic Medicine and Pathology* **18**:158–162.

Knight B (1996) The pathophysiology of death. In: *Forensic Pathology*, 2nd edn, pp 51–94. London: Edward Arnold.

Lawler W (1993) Bodies associated with fires. *Journal of Clinical Pathology* **46**:886–889.

Leadbeatter S and Knight B (1988) Resuscitation artefact. *Medicine Science and the Law* **28**: no. 3, 200–204.

Murphy GK (1991) The trials of Steven Truscott. *American Journal of Forensic Medicine and Pathology*, **12**:344–349.

Plueckhahn V and Cordner S (1991) *Ethics Legal Medicine and Forensic Pathology*, 2nd edn. Melbourne: Melbourne University Press.

Scene of Death

P Vanezis, Department of Forensic Medicine and Science, University of Glasgow, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0424

Introduction

When a body is discovered, one of the first considerations will be to decide whether the death is suspicious or overtly homicidal in nature or otherwise requiring special consideration such as a death in custody. In the majority of cases, the decision as to whether or not criminality is involved is reasonably straightforward. However, one should appreciate that there are many instances where the initial appearance is deceptive and it is therefore essential that all scenes be treated as suspicious until a full assessment is made.

Comprehensive examination of the place of discovery of the body – the scene – together with the collection of all relevant samples and their examination, will facilitate the investigator's quest to accurately establish the circumstances surrounding the death in question.

Definition

A scene of death is the place where a person has died. It should not be confused with where the body is discovered. It is not unusual for someone to have been

killed at one locus and then deposited elsewhere. It becomes further complicated by the fact that an actual incident leading to death, i.e. stabbing, may have occurred in an entirely different location from where death occurred. Most commonly, however, the place where the person is killed is the same as where the body is found (see **Table 1**).

Different Types of Scenes

The type of environment where the body is found will dictate the approach taken. Scenes may also be categorized by the type of death and/or method of disposal.

Indoors

Finding a body in a domestic environment is common and the vast majority of deaths are due to natural causes. Sadly many of those found are elderly or infirm people living alone and found some time after death. It is essential to establish the identity of the person and whether or not they lived in the house in which they are found. The room in which the deceased is found will also be important. For example, someone found collapsed in the toilet may well have died suddenly of a pulmonary thromboembolism or from gastrointestinal bleeding. The position of the body, considered together with the type of room, will frequently assist in assessing whether or not death is due to natural or unnatural causes. Disturbance in a room may well indicate that there has been some type of struggle and the death may, therefore, have resulted from this. Finding the body in an unusual location within premises, such as a garage or a cupboard will alert suspicion. One should also bear in mind that pets that are confined in rooms with their dead owner, as well as rodents and other scavengers, may eat away at the soft tissues.

Outdoors

Surface The usual method of discovery of a person found out of doors, e.g. in a field, is accidentally by a member of the public for example while exercising the dog. The question of whether the person has died at the scene or elsewhere and transported to the locus will need to be considered. The discovery of a body by the seashore will pose the question of whether the deceased had fallen into the sea and been washed ashore or fallen or in some other way and landed in the water from a vessel. Postmortem damage from animals and relocation of body parts may also sometimes be a feature of bodies found out of doors. Occasionally deaths due to hypothermia in outdoor environments, where there has been paradoxical undressing, will need careful consideration.

Table 1 The scene of discovery of the body in 634 cases

Scene of:				
Incident	*Death*	*Discovery*	*n*	*%*
Same	Same	Same	471	74
Different	Same	Same	123	20
Same	Same	Different	36	6

Burials and partially concealed remains In some instances the investigation begins with a search for a grave, where information is available regarding its possible existence. There are various techniques available, including the use of aerial photography and the occasional use of an infrared thermal camera, geophysical techniques (well established in the field of archaeology) and trained dogs. Bodies may be partially or fully buried in a number of situations which include: clandestine single graves, multiple graves containing a small number of individuals, e.g. from one family, and mass graves, e.g. those resulting from genocide. Recovery of remains in burials will require the combined efforts of the forensic team, with the forensic anthropologist, archeologist and pathologist all having a prominent role.

Water

Bodies recovered from water tend to decompose slower than those found in air except where the water is heavily polluted. Adipocere is a common feature of bodies found in damp conditions. The investigator will need to assess where the body had entered the water in relation to where the body was recovered.

Scene Management

Management of the scene comprises the following essential components:

1. Searching for a body (if applicable)
2. Discovery of a body and preliminary assessment
3. Appointing an investigating officer and forming a scene team, including relevant specialists
4. Briefing before and/or after visit to the scene
5. Health and safety considerations
6. Scene assessment and collection of evidence
7. Retrospective scene visit(s)

It is essential that the senior investigating officer forms a team to examine the scene. The members of the team will comprise the senior investigating officer or

equivalent (depending on jurisdiction), other relevant criminal investigation officers, uniformed officers to secure the scene, specialist scene of crime officer, photographer, various forensic scientists (types of specialties will depend on the type of scene), forensic pathologist, police surgeon and other helpers or experts depending on the circumstances.

The forensic team in assessing a scene will need to consider the following:

1. Location: whether dwelling house or other closed environment, outdoor site, burial, water etc.
2. Type of incident which will require specialist expertise: fires, explosions, firearms, industrial and transportation accidents, discovery of skeletal remains.
3. Number of deceased persons: most scenes involve one body. Other situations may involve serial killings (one or more scenes), a mass disaster or a mass grave.
4. Health and safety considerations: especially relevant in scenes involving fires; where there may be noxious substances; booby traps in explosion scenes; inaccessible and dangerous terrain; possible risk of infection.
5. Urgency of the examination in relation to site and circumstances: the location of the body may necessitate removal at the earliest opportunity; for example a body found by the seashore where the tidal movement may submerge the body or on a busy road or other environment where obstruction to public facilities for longer than necessary is undesirable.
6. Climatic conditions: the weather conditions at the scene, particularly if adverse, may dictate early removal to the mortuary or the erection of a tent for protection. Furthermore, early removal, and refrigeration, particularly in a hot environment is essential to prevent rapid decomposition.
7. Available resources and expertise: these will vary depending on various factors. Cost effectiveness can be achieved with good organization, careful preparation and good liaison between team members.
8. Security: it is important to establish the boundary of the scene, secure it and ensure that only authorized personnel are allowed within the boundary. All unauthorized persons must be kept away from a scene in order to prevent:

(a) Contamination of the scene
(b) The media from obtaining unauthorized information
(c) Interference from curious onlookers, souvenir hunters and others who may potentially disrupt the work of the forensic team.

Collection of Evidence

The task of the forensic investigator is to find traces left behind, such as fingerprints, bloodstains, fibers from clothing, dirt from shoes etc. and to locate matching materials on a suspect in order to provide objective evidence that they were present at the scene. The fundamental guiding principle which underlies the approach to any scene examination is attributable to Edmond Locard. He coined his 'theory of interchange' which in essence states; 'that the person or persons at the scene when a crime is committed will almost always leave something and take something away'. This doctrine of exchange or transfer is thus based on the observations that:

1. The perpetrator will take away traces of the victim and the scene;
2. The victim will retain traces of the perpetrator and may leave traces of himself on the perpetrator;
3. The perpetrator will leave behind traces of himself at the scene.

Documentation of the Scene

The scene must be documented in a systematic and thorough manner so that an accurate record is available for presentation as evidence. There are a number of ways in which the scene is documented.

'Stills' photography

It is essential that stills photography should be of the highest quality in order to provide an accurate permanent record of the scene. Photographs should be taken, working from the periphery of the scene through to the area where the body is sited. It is also necessary to photograph early before evidence is altered, e.g. footwear impressions, or before the body is moved. Photography can be taken of all the stages of the scene examination as necessary.

Notes, sketches and plans

These provide the examiner with an 'aide memoire' for assistance with the preparation of statements. They should be legible and should be signed and dated.

Video recording

This should cover the whole scene showing the relationship of surfaces and objects to each other. Although still images can be produced from video frames, quality is not as good as good quality conventional photography. The use of commentary at the time of the recording may also be used to explain visual images seen and be employed as an instructional device.

Computer reconstruction

Computerized aids may have a role to play in both presentation of evidence and reconstruction of the incident, e.g. in a vehicular accident. These are either site survey packages (incorporating computer aided design (CAD) systems) or portable plan drawing systems packages.

Role of Forensic Medical Practitioners at the Scene

Certifying death

Prior to the arrival of the pathologist, in the UK a police surgeon will attend and certify death. The doctor may also give the police his views as to criminality. However, this is not frequently possible because of the preference for the body not to be moved until the pathologist attends.

Initial attendance of the pathologist

When the forensic pathologist first arrives at the scene it is important to make an initial general assessment of the environment and the body within it. At this stage the pathologist will be guided by the senior investigating officer regarding which route should be taken to the body, bearing in mind the need to prevent contamination as well as safety considerations where applicable. The approach to the scene by the pathologist will vary according to the type of scene and the environment. For example retrieving a body from a shallow grave involves an entirely different strategy from dealing with the discovery of a fresh body in a public place. The pathologist will be appraised of the circumstances surrounding the discovery of the body and an initial view will usually be sought. However, although the pathologist may be able to give valuable guidance at an early stage one must exercise caution, mindful of the fact that at an early stage of any investigation, accounts of circumstances leading to death may be confused and inaccurate.

The position of body and its condition

It is crucial for the pathologist to carefully document the position of the body and to ascertain whether it had been moved and the reasons for doing so, prior to his/her attendance. There are a number of reasons why a body may be moved from its initial position. These include movement by someone, e.g. a relative, to see if the person is still alive; for resuscitation; to collect essential evidence that could be lost or contaminated; to secure the safety of the scene.

Examination of clothing

The presence and type of clothing as well as how it is worn and arranged on the body should be assessed. There may be various contaminants such as bloodstains, semen, saliva, paint smears, glass which needs to be looked for and collected where appropriate. The clothing may also show characteristic damage from cutting implements such as knives or from firearms or it may be torn during a fight for example. Such damage will yield valuable information which should be carefully assessed in conjunction with the body and its environment. For example the distribution of blood stains on clothing may assist the examiner to ascertain whether the victim was attacked while standing or lying down. Tears in clothing from stabbing will assist in assessing number of impacts and type and size of knife. In sexual assualt cases the clothing is frequently dishevelled, torn or some garments may be missing. It should be emphasized that a detailed examination of the clothing should be carried out in the autopsy suite to avoid undue disturbance of the body.

Associated signs such as bloodstains, drag marks various objects including furnishings

The immediate environment of the body and, where appropriate, other locations, for example a nearby room, should be examined to assess disturbance of furniture and other objects. This will allow an assessment to be made of the degree of struggle between two persons and whether any objects could have been used as weapons or whether the deceased may have impacted against furniture, floors etc. to cause the injuries seen. The presence of drag marks on floors caused by moving the body may be seen as smeared blood or dirt and there should be corresponding evidence on the body. The position of bloodstains, whether droplets, cast off stains from weapons or from arterial spurts should all be noted in relation to the position of the body.

Time of death assessment

Frequently, it is important to make an assessment of how long the person has been dead (postmortem interval: time of death). This is not an easy matter, but the pathologist can give useful guidelines as to the approximate time depending on the state of the body, the climatic conditions and the type of environment. Nevertheless, an estimation of time of death in certain circumstances may be both futile and misleading. This is particularly so if the body had been moved between different types of environments, e.g. indoor to outdoor, or been placed in cold storage before disposal, or been subjected to heat as in disposal by

fire. Details of how time of death is estimated is discussed elsewhere but it is important to consider the core body and ambient temperature, rigor mortis, hypostasis and general state of the body, including the effect of insect action when assessment is made. In any case where a rectal thermometer is inserted to obtain a core temperature, it is essential to remove and preserve the relevant undergarments and take all genital swabs prior to insertion of the thermometer. It should be appreciated that the ambient temperature may have fluctuated appreciably between when the body was first at the scene till its discovery. For example where a house is centrally heated, the temperature of the room in question may vary according to the timings at which the heating is set. An open window may also make a substantial difference to the temperature. It is also important to test for rigor mortis at the scene as this may disappear on arrival at the mortuary, particularly in hot climates where rigor is known to appear and disappear early. Hypostasis may also be useful in noting its extent, i.e. whether the body has been dead sufficiently long for it to be established. However, this phenomenon is of more use in assisting the pathologist in assessing whether the body has been in one position at the scene or whether at some stage prior to discovery, the body had been moved. Signs of putrefaction should also be carefully assessed as well as any insects found on the body as these findings will all be useful in allowing an overall assessment of the time of death to be made.

Taking samples

It is essential to take samples at the scene in a number of circumstances particularly where movement of the body or removal of clothing may cause contamination. Wherever possible, particularly in sexual assault cases it is recommended that the body is swabbed externally over the trunk, thighs and wherever else relevant, as well as taking swabs from the usual orifices. It is preferable, once the essential swabs have been taken, to leave nail clippings, hair samples and other samples to be taken in the autopsy suite.

Interpretation of the postmortem findings in relation to scene assessment

The pathologist should always conduct the autopsy bearing in mind the findings at the scene and whether injuries for example can be accounted for by the scene environment. In a number of cases it may be necessary to carry out a further visit to the scene in the light of autopsy findings. In cases where there was no opportunity to visit the scene where the incident was said to have occurred, e.g. in the vast majority of fatal 'battered child' cases, the pathologist may be requested to visit the scene and assess whether various explanations given by the parents or guardians are plausible.

Role of the Forensic Scientist and Scene of Crime Officer

The main concern of the scientist and scene-of-crime officer is the proper collection and preservation of evidence.

Care must be taken as to how the body is handled and moved from its site of initial discovery, as moving the body can lose or confuse evidence – particles fall off, blood swilling washes off materials etc. In sexual assault murders, moving the body can redistribute body fluids which can lead to erroneous conclusions. For example, semen can trickle into the anal passage of a victim and give the impression of attempted buggery.

When the body is not present, it can be difficult to interpret blood distribution patterns or assess particles from under or alongside the body. Therefore, it is always better for the 'team' to attend when the body is *in situ*. Discussion between the various team members will then allow decisions to be made on the order of examination. It is essential to remember that examining the body means entering, and disturbing the scene.

Examination of the body and clothing by the forensic scientist is essential for the identification of traces of marks which may have:

1. Come from an object, e.g. paint smears or glass particles left by a car involved in a fatal hit-and-run accident;
2. Been transferred to an object (Locard's principle, see above), e.g. blood, hair, fibers and fabric marks could be transferred from the body to the car;
3. Come from the person who carried out the attack, e.g. bloodstains, semen in rape cases, shoe prints found on the body or on the clothing and fibers transferred from the assailant's clothing. Particular attention should be directed to sites of attack and defense, e.g. fingernails, hands, forearms and face;
4. Come from another scene, e.g. paint fragments or fibers retained on the body which relate to a different venue from where body found.

Removal from the Scene to the Mortuary

Preparation for removal

One of the main difficulties encountered in preparing a body for removal from a scene is to prevent contamination from three main sources: the scene; one part of the body to another; all personnel handling the body. It is necessary therefore to place bags on the

hands, head and feet and to place the body in a bag with as little disturbance as possible and in the same orientation. For example a body that is found in a prone position should be placed in a bag in the same position. Any visible loose materials should have been collected from the body prior to removal. Where the body is in an advanced state of decomposition care should be taken that tissue or indeed parts of the body are not separated or modified in some way when the body is secured for removal.

Removal and transit to the mortuary

The removal and transit of the body from the scene to the mortuary may give rise to further problems if due care is not taken. Sometimes if the scene is in a confined space it may be difficult to remove the body without it coming into contact with surfaces such as stairs, walls or doors. The pathologist must therefore be aware of the possibility of postmortem injuries caused by removal in such circumstances.

Reception into the mortuary

Once the body arrives at the mortuary it is important to bear in mind that the body should remain in its bag until the pathologist starts the examination. It may be practical to place the body in a chill refrigerator at about 4°C until the pathologist arrives if there is likely to be an appreciable interval between reception at the mortuary and the autopsy. If time of death is an issue, then the body should be kept at room temperature for further core temperatures to be taken should these prove necessary.

See also: **Causes of Death**: Overview. **Crime-scene Investigation and Examination**: Recording; Recovery of Human Remains; Preservation. **Pathology**: Overview; Victim Recovery; Preservation of Evidence. **Photography and Digital Imaging**: Overview. **Time Since Death**.

Further Reading

Eckert WG and James SH (1989) *Interpretation of Bloodstain Evidence at Crime Scenes*. New York: Elsevier.
Fisher BAJ (1993) *Techniques of Crime Scene Investigation*, 5th edn. Boca Raton FL: CRC Press.
Gerberth VJ (1983) *Practical Homicide Investigation*. New York: Elsevier.
Haglund WD and Sorg MH (1997) *Forensic Taphonomy*. Boca Raton, FL: CRC Press.
Hunter J, Roberts C and Martin A (eds) (1996) *Studies in Crime: an Introduction to Forensic Archaeology*. London: BT Batsford.
Redsicker DR (1991) *The Practical Methodology of Forensic Photography*. New York: Elsevier.
Vanezis P and Busuttil A (eds) (1996) *Suspicious Death Scene Investigation*. London: Arnold.

Sharp Injury

D J Pounder, Department of Forensic Medicine, University of Dundee, Dundee, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0428

Introduction

Sharp force injuries are distinguished from blunt force injuries and represent the second major class of injuries. They result from contact with keen edges or sharp points. Generally sharp force injuries are subclassified as either incised wounds (cuts and slashes) or puncture wounds (stabs). An incised wound is a breach of the skin, of variable depth, resulting from contact with a keen edge. Usually there is at least penetration of the full thickness of the skin (i.e. both the epidermis and the dermis) together with variable involvement of the deeper tissues, but shallow, partial-thickness skin cuts can occur, for example a common 'paper cut'. Stab or puncture wounds are penetrating injuries with a depth greater than the wound dimensions on the skin surface. They are caused by long, thin objects, usually sharp-pointed, such as a knife.

Cuts and stabs may be assaultive, self-inflicted or accidental. The injury pattern assists in establishing the manner of infliction, but instances of single fatal stab wounds can be difficult to resolve. Incised wounds provide little information about the weapon, whereas stab wounds commonly permit at least a general reconstruction of weapon dimensions. Both incised wounds and stab wounds typically yield no trace evidence, but are associated with blood staining of the weapon, clothing and scene, opening up the possibility of crime scene reconstruction through bloodstain pattern interpretation. Questions which are often raised, but difficult to answer, include the degree of force required to inflict a stab wound, the order of infliction of the wounds, and the period of consciousness and level of activity possible following wounding.

Incised Wounds (Cuts)

The features of incised wounds reflect their production by keen-edged, rather than blunt, objects. Incised wounds have well-defined, cleanly severed margins in contrast to the ragged, torn edges of lacerations. Occasionally, incised wounds show some irregularity or notching of the wound margins resulting from cutting across skin folds, a feature most commonly seen in defensive injuries to the palms (see below). Typically incised wounds have no associated abrasion

or bruising of the adjacent skin, and there is complete severance of all tissues within the depths of the wound, so that there are none of the 'tissue bridges' which characterize lacerations. The distinction between incised wounds and lacerations is of considerable forensic importance. The diagnosis of an incised wound implies contact with a sharp-edged object or weapon, whereas the diagnosis of a laceration implies contact with a blunt object or weapon as the cause of the injury. Unfortunately, in clinical medicine, this clear forensic terminology is often not maintained, so that incised wounds and lacerations are both loosely referred to as lacerations. This may create confusion when medical records are used in legal proceedings.

Although the presence of an incised wound indicates that the causative object was keen-edged, the wound commonly provides no further information about the weapon. Most incised wounds are knife wounds, but this is not a conclusion that can be safely drawn from the examination of the wound alone. Cuts may also be produced by razor blades, shards of glass, sharp edges of machinery and tools, tin cans and similar objects. Whatever the nature of the keen-edged object producing the incised wound, it cleanly severs the tissues and typically leaves no trace evidence within the wound. However, the object will have traces of blood and tissue, which can be used to link a victim and a weapon through DNA fingerprinting. Cleanly severed blood vessels within the wound bleed profusely, and if large vessels are cut then the hemorrhage may be sufficient to kill. Such extensive hemorrhage commonly results in blood soaking of clothing together with blood staining and spattering of surroundings, opening up the possibility of reconstruction of the events through blood spatter analysis. When large veins are severed, particularly in the neck, air may be drawn into the circulation, obstructing the flow of blood through the heart and so killing by air embolism. The diagnosis of air embolism is easily missed at autopsy if it is not considered and actively sought at the outset.

Incised wounds inflicted with heavy-bladed weapons, such as an axe, adze, sharpened spade or machete, or by a boat propeller, characteristically show abrasion, and sometimes bruising, of the wound margin. This results when the edges of the first-formed incised wound are crushed by the forceful entry of the heavy blade following behind. The extent of the abrasion is highly variable and is sometimes only apparent on close inspection, perhaps aided by a hand-lens. However, the large size of the wounds and the commonly associated cuts to bone are more obvious pointers to the type of weapon. These wounds are commonly described as chop wounds.

Accidents

Accidentally inflicted incised wounds are common. They may result from the use of knives in the home or workplace, broken window glass (in burglaries), broken drinks glasses, and sharp-edged machinery or parts of vehicles. The commonest accidental incised wounds are probably the simple, shallow cuts produced by stiff paper or tough grasses.

Frequently seen is the pattern of multiple, incised wounds to the backs of the hands and to the face, caused by the shattering of the tempered-glass side windows of vehicles during collisions. The resultant dice-like fragments of glass produce the characteristic scattered, small, irregular incised wounds with associated abrasions. They involve the side of the body adjacent to the shattered window, so that it is possible to determine whether the individual was seated on the left or right side of the vehicle at the time of impact. In this way the injuries can be used to distinguish the driver from the front-seat passenger.

Self-infliction

The ready availability of knives and the relative painlessness of incised wounds have encouraged their use in suicide since antiquity. The Romans were said to have favored opening the veins of the arms whilst lying in a warm bath. Suicide by cutting the throat was far commoner in the years before the wide availability of relatively painless drugs and poisons, so that the best descriptions and illustrations are in the older texts. Suicidal incised wounds are typically directed towards sites, such as the neck, wrists and, less commonly, the elbows, knees and ankles, where large blood vessels are close to the skin surface. The cuts are made by first pressing the sharp edge against the skin and then running it across the skin surface, so that the shallower wounds are both of uniform depth and respect the general contours of the body. Self-inflicted incised wounds are commonly, but not invariably, accompanied by multiple, parallel, and sometimes superimposed shallow wounds, known as hesitation or tentative wounds. Such wounds reflect a testing of the blade as well as the indecision so often present in suicidal acts. Having produced an incised wound, the suicide may then repeat the cuts in the depths of the first-formed wound. This behavior is inferred from the multiple trailing cuts arising from the ends and sides of the principal wound. Occasionally these repeated cuts make the wound extremely deep, and death investigators may find it difficult to accept the possibility of self-infliction. For example, in suicide by cutting the throat there may be repeated cut marks on the exposed cervical spine associated with severing of all the anterior neck structures.

Parallel, shallowly incised wounds to the wrists and forearms, resulting from self-harm, may heal leaving multiple linear scars. Often they are most prominent on the nondominant arm and involve the extensor and flexor surfaces of the forearms and the volar surface of the wrists. If seen in clinical forensic practice these scars raise the possibility of a personality or psychiatric disorder, with the attendant risk of suicide in custody. If seen at autopsy, they raise the index of suspicion for suicide, alcohol and drug abuse, and risk-taking behavior leading to accidental death.

Self-inflicted cuts are sometimes seen in individuals who falsely allege that they have been assaulted. The motivation is most often the manipulation of personal relationships at times of stress, but may involve financial gain, such as claiming state compensation for victims of assault. The pattern of injury involves accessible sites, and reflects handedness, infliction by pressure rather than slashing, and the avoidance of sensitive or critical areas such as nipples, lips and eyes. Even so, the injuries can be extensive and mutilating.

Assaults

Defensive injuries to the hands and forearms are typical of knife attacks. They may be absent in sudden, overwhelming attacks (so called 'blitz' attacks), or if the victim is unable to offer a defense as a result of the effects of alcohol and drugs, unconsciousness, bindings, or other physical and emotional circumstances. Since defensive wounds reflect anticipation of injury and an attempt to ward off the harm, they may be seen in accidents as well as in assaults. For example, multiple cuts to the palms of the hands can be produced in a fall onto a glass-strewn floor, as well as during an attack with a broken bottle. In a knife attack, defensive cuts to the palms and fingers result from attempts to grab or to deflect the weapon, whereas slash and stab wounds to the backs of the hands and the forearms result from shielding movements. A victim on the ground, being attacked by a standing assailant, may have defensive wounds to the legs. In cases of suicide, typical defensive-type wounds to the hands and arms do not occur, but the occasional shallow cut to the fingertip may reflect testing the keenness of the blade.

Assaultive incised wounds inflicted on a resisting victim are typically scattered, include defensive-type wounds and involve the clothing, which, by contrast, is characteristically undamaged in self-infliction. Wounds to areas of the body that are difficult to access for self-infliction, e.g. the back, immediately raise the suspicion of assault. Although it is rarely, if ever, possible to determine the precise order of infliction of the assaultive wounds, as a general rule the scattered injuries are inflicted first during the phase of active resistance, whereas the closely grouped injuries, directed towards vital target sites, are inflicted later. In knife assaults it is usual for there to be both incised wounds and stab wounds, with death usually the result of the deeply penetrating stab wounds rather than the shallower cuts. A combination of incised and stab wounds may also be seen in suicides.

Stab Wounds

Stab wounds are penetrating injuries produced by a long, thin object, which is typically pointed. Most commonly, the instrument is flat with a sharp point, such as a knife, a shard of glass or a length of metal or wood. Other weapons may be long and thin with a sharp point, such as a skewer, hypodermic needle, ice pick, or old-fashioned hatpin. With sufficient force even long, rigid objects that are blunt-ended will produce puncture wounds, e.g. screwdriver, wooden stake or protruding parts of machinery or motor vehicles. The appearance and dimensions of the resulting wound often provides useful information about the object producing it. Stab wounds, like incised wounds, typically yield no trace evidence, are associated with blood and tissue staining of the weapon, as well as blood staining and damage to clothing, and offer the opportunity for scene reconstruction from blood spatter analysis.

Clothing

Observations and reports on stab wounds should always attempt to correlate damage to the clothing with wounds to the body, because this very much facilitates reconstruction of the incident. Aligning cut and stab holes to the clothing with cut and stab wounds to the body may help determine clothing position and body position at the time of stabbing. Some stabs or cuts to clothing may not penetrate to produce wounds, or there may be multiple holes in the clothing reflecting a single stab through folds. Some wounds may not be associated with damage to all layers of the clothing, for example sparing a coat or jacket. Suicides typically, but not universally, adjust their clothing to expose the bare skin before inflicting the injuries, so that the clothing is disturbed but undamaged. Any bloodstain pattern to the clothing is best interpreted in the light of the wounds to the body and the scene examination.

Wound description

If there are multiple wounds it is often useful to number them. Each wound should then be described giving details of its site relative to local anatomical

landmarks as well as its distance from the midline and distance above the heel. Next the shape, size and orientation of the wound should be recorded and the dimensions assessed with the wound edges reapposed. Unusual marks to the surrounding skin, such as bruises and abrasions need to be described since, however small, they are likely to prove significant. The wound should never be probed with a sharp object, and particularly not with an alleged weapon, but rather the direction should be assessed during dissection. Oblique penetration of the blade through the skin may leave one wound edge bevelled and the opposite edge overhanging, i.e. undercut, giving an indication of the direction of the underlying wound track. The depth of the wound track needs to be estimated, recognizing that it is subject to some considerable error. Damage to tissues and organs is described sequentially and thereafter the effects of the damage, e.g. hemorrhage, pneumothorax, air embolism, is recorded. A wound which passes completely through a structure is described as perforating or transfixing or 'passing through and through'. If the wound track enters a structure but does not exit it then the wound is described as penetrating. Following this convention, a stab wound which passed through the front of the chest to end in the spinal column could be described as a penetrating stab wound of the chest which perforated the sternum, perforated the heart and penetrated the vertebral column. Systematic observation and recording of the site, shape and size, direction, depth, damage and effects of each wound is the hallmark of good practice. Photodocumentation of the wounds with appropriate measuring scales is standard in suspicious death investigation, and advisable in any potentially contentious case.

The skin surface appearance of a stab wound is influenced both by the nature of the weapon and by the characteristics of the skin. The skin contains a large amount of elastic tissue which will both stretch and recoil. This elastic tissue is not randomly distributed but is aligned so as to produce natural lines of tension (Langer's lines) which have been mapped out on the skin surface, and are illustrated in standard anatomy texts. The extent of wound gaping, and the extent of wound scarring in survivors of knife assaults, is influenced by the alignment of the wounds relative to Langer's lines. Stab and incised wounds which are aligned with their long axis parallel with Langer's lines gape only slightly, a fact made use of by surgeons who align their incisions in this way to promote healing and reduce scarring. Wounds aligned at right angles to Langer's lines tend to gape widely, and scar prominently, because the natural lines of tension of the skin pull the wound open. Wound gaping is also influenced by the extent of damage to the underlying supporting fascia and muscles. Where a stab wound is gaping then the wound edges must be re-approximated at autopsy to reconstruct the original shape of the wound, something that is easily achieved with transparent tape. It is the dimensions of the reconstructed wound rather than the original gaping wound that are of interpretative value.

Weapon dimensions

If a stabbing with a knife is 'straight in and out' then the length of the stab wound on the skin surface will reflect the width of the knife blade. There are important qualifications that apply to this however. The skin wound length may be marginally (a few millimeters) shorter than the blade width as a result of the elastic recoil of the skin. If the knife blade has a marked taper and the entire length of the blade did not enter the body then the skin wound length will not represent the maximum width of the blade. If the blade did not pass 'straight in and out' but instead there was some 'rocking' of the blade, or if it was withdrawn at a different angle from the original thrust, then the skin wound will be longer than the inserted blade width. Consequently, the most reliable assessment of blade width is made from the deepest wound with the shortest skin surface length. A single weapon can produce a series of wounds encompassing a wide range of skin-surface lengths and wound depths. This is often seen in a multiple stabbing fatality and is consistent with the use of only one weapon. However, it is rarely possible to exclude any speculative suggestion of more than one weapon, and, by inference, more than one assailant.

The depth of the wound gives an indication of the length of the weapon. Clearly the wound track length may be less than the blade length if the entire blade did not enter the body. Less obvious is the fact that the wound track length may be greater than the blade length. This occurs if the knife thrust is forceful and the tissues are compressed, so that when the weapon is withdrawn the track length in the now decompressed tissues is greater than the blade length. This tissue compression effect is most marked in wounds to the anterior chest and abdomen, since the entire chest or abdominal wall can be driven backwards by the blow. A small pocket-knife, with a blade less than 5 cm (2 in) can cause, in a slim person, a fatal stab wound to the heart or one which penetrates the abdomen to transfix the aorta. An added complication in measuring the wound track length at autopsy is that the corpse is supine with the viscera in a slightly different relative position to a living person standing or sitting. For all of these reasons the wound

track depth should be used with caution in predicting the blade length of the weapon. If by chance some fixed bone, such as a vertebra, is damaged at the end of the wound track, then the assessment of depth of penetration is easier, but still subject to inaccuracy.

As well as providing an indication of blade width and length, a stab wound may provide other useful information about the weapon. Wound breadth on the skin surface is a reflection of blade thickness and a typical small kitchen knife, with a blade thickness of 2 mm or less, produces a very narrow wound. The use of a thicker-bladed weapon may be readily apparent from the measured wound breadth on the skin surface. Most knives have a single-edged blade, that is one keen edge and one blunt edge to the blade. The resultant wound reflects the cross-sectional shape of the blade and, with the wound gaping, often appears boat-like with a pointed prow and a blunted stern. Sometimes the blunted stern shape is distorted into a double-pronged fishtail. The thicker the blade of the weapon, the more obvious is the blunting of one end of the wound when contrasted with the other pointed end. Knives with double-edged blades (daggers) are specifically designed for use as weapons and produce a wound that is pointed at both ends, but such a wound may not be distinguishable from one produced by a thin, single-edged blade. The cross-sectional shape of the blade of the weapon may be accurately reproduced if it passes through bone, e.g. skull, pelvis, sternum or ribs. At the same time, trace material, such as paint, present on the blade will be scraped off by the bone and deposited in the wound. Stab wounds in solid organs such as the liver may retain the profile of the weapon, and this can be visualized by filling the wound track with a radio-opaque contrast material and taking a radiograph.

Stab wounds inflicted during a struggle, with knife thrusts at awkward angles and with movements of both parties, may show characteristics reflecting this. Even so, it is rarely if ever possible to reconstruct the positions of victim and assailant from the location and direction of the wounds. A notch on the otherwise cleanly cut edge of the wound is a result of withdrawal of the blade at a different angle from the entry thrust. Exaggeration of this effect leads to a V-shaped, or even cruciate, wound when there is marked twisting of the blade or twisting of the body of the victim. A linear abrasion (scratch), extending from one end of the wound, results from the withdrawing blade tip running across the skin. A single stab hole on the skin surface may be associated with more than one wound track through the tissues, reflecting a knife thrust followed by continuing struggle or repeated thrusts of the weapon without complete withdrawal. If the entire knife blade is forcefully driven into the body then there may be bruising or abrasion of the surrounding skin from the hilt of the weapon. More commonly, there is a small rectangular abrasion at one end of the wound reflecting penetration of the blade beyond the notch at the base of the blade (the 'kick') which separates the sharpened edge from the rectangular portion of metal from which the blade was forged (the 'tang'). Consequently the presence of this small abrasion is an indication that the blade penetrated to its full length.

Weapons other than knives may produce characteristic stab wounds. Bayonets, which have a ridge along the back of the blade and a groove along either side, to facilitate withdrawal of the weapon, may produce a wound like an elongated letter 'T'. A pointed metal bar which is square in cross-section typically produces a cruciate wound, whereas one which is circular in cross-section, e.g. a pitchfork, produces an elliptical wound. A closed scissors produces a compressed Z-shaped wound, and a triangular file will produce a three-cornered wound. If the cross-sectional shape of the weapon varies along its length, e.g. a screwdriver, then the depth of penetration will affect the appearance of the wound. Relatively blunt instruments such as pokers, closed scissors and files, tend to bruise and abrade the wound margins, a feature not otherwise seen in stab wounds. The blunter the object and the thicker its shaft then the more likely is the skin surface wound to become a ragged, often cruciate, split. In cases where the wound appearance is unusual, it is helpful to conduct experiments with a duplicate suspect weapon in order to see whether the appearance of the wound can be reproduced.

Degree of force

A commonly asked question in the courts is the amount of force required to produce a specific stab wound. This is usually a difficult if not impossible question to answer. The sharpness of the point of the weapon is the most critical factor in determining the degree of force required to produce a stab wound. In general, relatively little force is required to produce a deeply penetrating stab wound using a sharply pointed weapon, and the amount of force is easily overestimated. The greatest resistance to penetration is provided by the skin and once this resistance is overcome the blade enters the tissues with greater ease. In this respect an analogy can be made with the stabbing of a ripe melon. The important implication is that the depth of the wound is not a measure of the degree of force applied. However, penetration of any bone or cartilage implies a significant degree of force, all the more so if the tip of the blade has broken off and remains embedded in the bone, something which is

best identified by X-ray. Similarly a significant degree of force may be inferred from the presence of the hilt mark of the weapon on the skin surface, an uncommon finding, or a wound track significantly longer than blade length, suggesting forceful tissue compression during the stabbing (see above). Even so, the stabbing force may have been a combination of both the thrust of the weapon and also any forward movement of the victim, such as in a fall. This latter proposition is commonly raised by the defense, and is rarely possible to discount, in deaths from single stabs.

Homicide, suicide and accident

Most stabbing deaths are homicidal and the wounds are usually multiple, since the first wounds will usually leave the victim capable of some continuing resistance. The stab wounds are commonly associated with incised wounds and the typical pattern is of scattered wounds, many deeply penetrating, with wound tracks in different directions and several potentially lethal. Wounds to the back and other sites inaccessible to the victim all suggest homicide, as do a large number of scattered wounds. Defensive wounds to the hands and arms are common but their absence does not exclude homicide. Multiple grouped shallow wounds in a homicide may reflect threatening or sadistic behavior prior to the lethal blows, or alternatively postmortem piquerism.

In suicidal stabbings the wounds are usually multiple and closely grouped over the left breast or upper abdomen with no associated damage to the clothing which is characteristically removed or pulled aside, although this is not invariable. Occasionally suicidal stabs are inflicted on the neck. Some wounds are shallow representing tentative thrusts, and only a few are deeply penetrating and potentially lethal. Multiple wound tracks arising from a single stab hole reflect repeated thrusts following partial withdrawal of the blade. There may be associated incised wounds to the wrists and neck. A small percentage of suicides will use more than one method, so that a drug overdose is followed up with cutting and stabbing, or the knife wounds followed by hanging or drowning.

The most difficult cases are those with a single stab wound and no other injuries. Allegations of homicide are usually met with a defense of accident. In other circumstances distinguishing suicide from homicide may be difficult. Some suicides may accomplish the act with a single thrust of the weapon which is left protruding from the wound. Very occasionally the hand of the decedent may be gripping the weapon, in cadaveric spasm. Although this is proof that the victim was holding the weapon at the time of death, it is not conclusive proof of suicide.

Collapse and death

A lethal injury can lie beneath the most trivial looking skin wound, emphasizing the fact that the primary characteristic of a stab wound is its depth. Stab wounds kill mainly as a result of hemorrhage, which is usually internal rather than external. Indeed, there may be little bleeding onto the skin and clothing in a fatal stabbing. Whereas the loss of a large volume of blood will clearly account for death, a much smaller blood loss, of about one liter, may have a fatal outcome if it occurs very rapidly, as in cases with gaping wounds to the heart or aorta. As happens with incised wounds, death can also result from air embolism, in which case collapse usually occurs very rapidly. Both air embolism and pneumothorax are possibilities to be explored at the start of the autopsy. Those victims who survive the immediate trauma of a stabbing may succumb to infections after many days or weeks.

Initially the victim of a stabbing assault may be unaware of having received a wound. The adrenaline-surge, preoccupation with a struggle, intoxication with alcohol or drugs, the relative painlessness of sharp force injuries, and the tendency of stab wounds to bleed internally rather than externally, are some of the factors contributing to this lack of awareness. Consciousness and physical activity can be maintained for a variable period after the stabbing, depending on the rate of blood loss or the development of air embolism. Even after examining the wound it is difficult to give a useful estimate. Following a stab wound to the heart, at least a few minutes of activity and consciousness is common. Stabs to the brain can produce immediate incapacity, or even death, depending on the part damaged, and similarly stab wounds severing large nerves of the limbs can immediately paralyse that limb. Otherwise incapacitation reflects the rapidity and the extent of blood loss. Since the development of air embolism can be delayed, its presence, although typically associated with sudden collapse, is not necessarily an indicator of instant collapse.

See also: **Clinical Forensic Medicine**: Defense Wounds; Self-inflicted Injury.

Further Reading

DiMaio VJM and Dana SE (1998) *Handbook of Forensic Pathology*. New York: RG Landes.

DiMaio DJ and DiMaio VJM (1989) *Forensic Pathology*. New York: Elsevier.

Knight B (1996) *Forensic Pathology*, 2nd edn. London: Edward Arnold.

Tedeschi CG, Eckert WG and Tedeschi LG (eds) (1977) *Forensic Medicine*. Philadelphia: Saunders.

Sudden Natural Death

M Burke, Victorian Institute of Forensic Medicine, Australia

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0426

Introduction

In addition to performing a postmortem examination on the body of an individual who dies as a result of trauma, forensic pathologists may also examine the body of individuals who have died as a consequence of natural disease processes.

Sudden natural death as investigated by a coroner or medical examiner may have significant medicolegal ramifications in addition to providing valuable information to the deceased's immediate family and the community as a whole.

Definition

The definition of sudden natural death varies between organizations and jurisdictions. Although the World Health Organization defines sudden death as occurring within 24 h of the onset of symptoms, many forensic pathologists would only include cases where the death has occurred within hours of signs and symptoms or where the immediate clinical history is unknown.

Common Causes of Sudden Natural Death

Cardiovascular disease and, in particular, coronary artery atherosclerosis, is the most common cause of death in Western societies.

The resultant myocardium ischemia caused by coronary artery atherosclerosis may lead to a lethal cardiac arrhythmia and be the first sign of occult disease.

Hypertensive heart disease in the presence of normal coronary arteries can lead to sudden death as a consequence of a cardiac arrhythmia. Hypertension can contribute to sudden death by precipitating rupture of an abdominal aortic aneurysm or dissection of the thoracic aorta, or in causing an intracerebral hemorrhage.

Pulmonary thromboembolism is a relatively common cause of sudden death with most cases associated with the predisposing factors of trauma and immobility.

The following discussion examines the varied causes of sudden natural death with respect to the individual organ systems. The order of the systems examined and the description of each disease process within a particular system, is a reflection of the relative incidences of that disease.

Systems

Cardiovascular system

Coronary artery atherosclerosis Coronary artery atherosclerosis is the most common cause of sudden death in Western societies. The stenosis or occlusion of a particular coronary artery results in ischemia of the supplied region of myocardium which may result in acute myocardial infarction and/or a lethal cardiac arrhythmia. Risk factors for coronary artery atherosclerosis include smoking, hypertension, diabetes mellitus, hyperlipidemias, obesity and familial factors.

Coronary artery atherosclerosis tends to occur in the proximal aspects of the left main coronary artery, right main coronary artery, at branching points and the proximal aspects of the left anterior artery and left circumflex arteries. Acute critical occlusion of a vessel may result in the sudden onset of symptoms. The acute change may result from complications within an atheromatous plaque including plaque rupture with subsequent thrombosis of the vessel and sudden hemorrhage into a plaque resulting in loss of luminal area and subsequent myocardial ischemia.

As most forensic cases of sudden natural death due to coronary artery atherosclerosis occur within hours of the onset of symptoms, the pathologist does not usually find acute myocardial infarction upon macroscopic and microscopic examination of the heart. The earliest features of acute myocardial infarction include interfiber edema and contraction band necrosis, findings which typically require hours of survival postinfarction but which can be seen as an artifact of resuscitation.

It is generally accepted that at least 12–24 h of survival postinfarction must occur for the earliest recognizable histological changes to evolve in the heart. One may see eosinophilia of myocardial fibers with loss of nuclear detail followed by neutrophil infiltration, often peaking at about 24–36 h survival postinfarction. Enzyme histochemistry and immunoperoxidase stains have been used in an attempt to reveal earlier stages of infarction, however, it would appear that these special stains are best demonstrated when infarction is easily seen on routine hematoxylin and eosin-stained sections.

Organization or repair of an acute myocardial infarct is characterized by macrophages removing

necrotic tissue and new blood vessel formation followed by collagen deposition by fibroblasts. Important early complications of acute myocardial infarction include cardiac arrhythmias and congestive cardiac failure. Later complications include ventricular wall rupture, mural thrombosis with embolization, aneurysm formation and cardiac failure.

Hypertensive heart disease Hypertensive heart disease may be reliably diagnosed from the pathological features of an enlarged heart showing concentric left ventricular hypertrophy in the absence of valve disease or cardiomyopathy. Objective evidence for cardiomegaly may be obtained from referring to published studies of heart weight with respect to sex, height, weight and body size. Characteristic hypertensive changes may be observed in the kidney and in older individuals, the lenticulostriate vessels within the basal ganglia of the brain.

Pulmonary thromboembolism Pulmonary thromboembolism is a common cause of sudden death especially in individuals with a recent history of trauma and/or immobility. Peripheral venous thrombosis occurs as a result of (1) status of blood flow, (2) damage to the endothelial lining of the blood vessel and (3) abnormalities in blood coagulation. Recent studies have indicated that abnormalities in coagulation factors may have a familial basis with obvious ramifications for immediate family members.

Aortic aneurysm Rupture of an abdominal aortic aneurysm results in sudden collapse from massive exsanguination. Typically an elderly individual, often with the typical risk factors for coronary artery atherosclerosis, complains of severe abdominal pain culminating in collapse from catastrophic hemorrhage. Dissection of the thoracic aorta may cause death as a consequence of rupture into the pericardial sac with subsequent pericardial tamponade, rupture into the pleural cavity resulting in exsanguination or from dissection of major arteries such as the coronary and carotid arteries with associated organ infarction. Aortic dissection may occur in individuals with hypertension and in individuals with genetic disorders of collagen formation such as Marfan's and Ehlers–Danlos syndrome.

Valve disease Valve disease is a not uncommon cause of sudden death in elderly individuals. The relative decline in the incidence of rheumatic heart disease has resulted in relatively more cases of congenital and degenerative valve disease with dystrophic calcification. Stenotic or incompetent valves lead to compensatory physiologic hypertrophy or dilatation of the heart muscle in order to maintain cardiac output which, however, will eventually result in heart failure and the possibility of lethal cardiac arrhythmias.

The following entities may also occur in older persons, but their relationship to sudden death in young fit individuals belies the rarity of the conditions.

Myocarditis refers to inflammation within the myocardium associated with fibre necrosis. The myocardial inflammation may be a direct result of a viral infection or from the effects of stimulation of the immune system. Depending on the degree of muscle involvement, an individual may suffer with symptoms of cardiac failure or may die as a consequence of a lethal cardiac arrhythmia.

A cardiomyopathy refers to a disorder of heart muscle. A classification of the cardiomyopathies according to the World Health Organization includes (1) dilated cardiomyopathy, (2) hypertrophic cardiomyopathy, (3) restrictive cardiomyopathy and (4) arrythmogenic right ventricular dysplasia. It is believed that some cases of dilated cardiomyopathy may occur as a consequence of the effects of a remote viral myocarditis whereas other etiological factors include exposure to toxins or familial disorders. Hypertrophic cardiomyopathy is a genetic disorder characterized by disarray of myocardial fibers on microscopic examination of heart muscle and may be associated with sudden death. Arrythmogenic right ventricular dysplasia shows inflammatory and fibrofatty infiltration of the right ventricle associated with ventricular tachyarrhythmias and is believed to be a cause of sudden death.

In rare cases no morphological abnormality or uncertain anatomical abnormalities are identified at postmortem in individuals thought to have died of a sudden cardiac event. In some of these cases conduction abnormalities such as the long QT syndrome may be responsible for the death. Such cardiac arrhythmias may occur as a consequence of a functional abnormality or may be due to the effects of drugs. Obviously as no definite morphological abnormality can be demonstrated the pathologist is heavily dependent on supportive clinical information in formulating a diagnosis.

Nervous system

Sudden and unexpected death attributable to the central nervous system most commonly occurs as a consequence of intracranial hemorrhage. Individuals with hypertension may succumb to hemorrhages within the basal ganglia, pons or cerebellum. The presence of the resultant hematoma can lead to death from a critical rise in intracranial pressure

with obstruction to the flow of cerebrospinal fluid or, especially in the case of pontine hemorrhages, may result in lethal brainstem dysfunction.

Hemorrhage into the subarachnoid space most commonly occurs from rupture of an aneurysm within the circle of Willis at the base of the brain. Less commonly the hemorrhage can be due to rupture of arteriovenous or other vascular malformation. Death is believed to occur due to the volume of blood within the subarachnoid space in addition to autonomic instability.

Forensic pathologists and clinical neurologists are aware of deaths in individuals with epilepsy in the absence of morphological abnormalities. Sudden unexpected death in epilepsy (SUDEP) is not associated with status epilepticus and is believed to relate to cardiac arrhythmias which have been demonstrated to occur during epileptiform seizures.

Although most cases of central nervous system tumors will present with localizing or general symptoms and signs prior to death, in some instances acute hemorrhage into the tumor or other local complications may lead to sudden death with the diagnosis made at autopsy. A rare cause of sudden death due to central nervous system disease is the colloid cyst of the third ventricle. In this condition a thin-walled cyst within the third ventricle causes acute obstruction to cerebrospinal fluid flow with subsequent acute hydrocephalus and death.

Sporadic cases of meningoencephalitis are most commonly associated with *Neisseria meningitidis* with septicemia and adrenal hemorrhage.

Gastrointestinal system

Sudden death in disease processes occurring in the gastrointestinal system primarily involve massive gastrointestinal hemorrhage. The etiological factors are protean and include esophageal varices, peptic erosions and ulceration and tumors.

Individuals with cirrhosis of the liver from any cause may develop varices of submucosal esophageal veins which are prone to trauma and subsequent massive hemorrhage. The concomitant liver dysfunction results in coagulation abnormalities which exacerbates the hemorrhage.

Peptic ulceration is often associated with the bacterium *Helicobacter pylori* and may cause torrential hemorrhage if erosion of a submucosal artery occurs. Widespread gastric erosions are seen in severe ‘stress’ of any cause and is believed to relate to increased circulatory glucocorticoids. Uncommonly, benign and malignant upper gastrointestinal hemorrhages may cause fatal hemorrhage.

Acute and chronic pancreatitis is most commonly related to gallstones or excessive alcohol intake but is also seen in hypertriglyceridemia, the vasculitides and other inflammatory conditions. The severe end of the spectrum of inflammation seen within the pancreas, hemorrhagic pancreatitis, can lead to death due to multiple system failure induced by systemic inflammatory mediators in the systemic inflammatory response syndrome.

In cases where medical attention is not sought or in rare cases where the diagnosis is unrecognized, acute surgical emergencies such as acute appendicitis, perforated peptic ulcer, obstructed bowel, mesenteric thrombosis and peritonitis may present as a sudden unexpected death.

Respiratory system

Sudden death is described in individuals with asthma. Analogous to the syndrome of sudden death in epilepsy, individuals with asthma may suffer sudden and unexpected death and not necessarily with a clinical acute asthmatic attack. Although the mechanism is obscure, it is believed to relate to a lethal cardiac arrhythmia.

Acute epiglottitis is an acute infective condition of the epiglottis that is well-recognized in children but which may also occur in adults. The causative organism is usually *Haemophilus influenzae*. Death results from acute upper airway obstruction.

Pneumonia may lead to ‘sudden death’ in socially isolated individuals and in other cases where the gravity of the infection is not recognized. The common community-acquired causative organism is *Streptococcus pneumoniae*. Rare cases of Legionnaire’s disease may be seen.

Occult malignant tumors may erode major blood vessels resulting in torrential hemorrhage with exsanguination and airway obstruction.

Endocrine system

Diabetes mellitus can lead to sudden unexpected death as a consequence of diabetic ketoacidosis. Often, in retrospect, a young individual may have had a history of polydipsia and polyuria and been vaguely unwell before death ensued. Postmortem toxicological examination can reveal a raised glucose concentration within the vitreous humor of the eye. Microscopic examination of the kidney may show vacuolization of the collecting ducts.

Genitourinary system

Sudden deaths in relation to the kidneys and urinary system are relatively rare. An individual may have underlying chronic renal impairment which may

deteriorate rapidly from a further insult such as urinary tract infection, dehydration or administration of a nephrotoxic drug. The diagnosis requires collation of all clinical, pathological and toxicological data.

Sudden death may occur as a complication of pregnancy. An ectopic pregnancy usually occurring in the Fallopian tube may rupture causing death from exsanguination from massive intraperitoneal hemorrhage.

See also: **Causes of Death**: Overview.

Further Reading

Hirsch CS and Adams VI (1993) Sudden and unexpected death from natural causes in adults. In: Spitz WU *Medico-legal Investigation of Death. Guidelines for the Application of Pathology to Crime Investigation*, 3rd edn, pp 137–174. Springfield, IL: Charles C Thomas.

Knight B (1996) The pathology of sudden death. In: Knight B (ed.) *Forensic Pathology*. London: Oxford University Press.

Systemic Response to Trauma

M Burke, Victorian Institute of Forensic Pathology, Southbank, Victoria, Australia

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0433

Introduction

The common causes of death from the local effects of trauma include exsanguination from the disruption of major blood vessels or organs, severe head injury, airway obstruction, and respiratory insufficiency from hemopneumothorax or mechanical dysfunction.

Individuals who initially survive an episode of trauma by way of the resuscitative actions of medical and surgical intervention may later suffer from the adverse systemic effects of trauma. Adverse systemic effects of trauma include the systemic inflammatory response syndrome (SIRS), adult respiratory distress syndrome (ARDS), disseminated intravascular coagulation (DIC), systemic fat embolism and venous thrombosis with pulmonary thromboembolism.

The SIRS refers to the effects of the systemic actions of the various mediators of acute inflammation. SIRS may progress to more severe forms where organ systems fail to maintain homeostasis resulting in multiple organ dysfunction syndrome (MODS). A continuum exits to the most severe end of the spectrum, i.e. multiple organ failure syndrome (MOFS).

ARDS refers to the acute lung injury that may complicate trauma and includes cases where there is no direct pulmonary trauma. Analogous to the adverse effects of inflammatory mediators in SIRS, the underlying pathogenesis of ARDS is a complex interplay between diverse pro- and anti-inflammatory mediators.

Disseminated intravascular coagulation describes generalized microthrombus formation in the systemic circulation. DIC occurs as a consequence of the failure of homeostasis within the coagulation system with massive thrombin generation and activation. DIC is involved in the pathogenesis of SIRS and is closely linked to the development of MODS.

Systemic fat embolism refers to the presence of globules of fat within the systemic circulation leading to organ dysfunction. Systemic fat embolization may occur following ‘saturation’ of the pulmonary circulation with spill over into the systemic circulation and/or the coalescence of unstable lipoproteins present within plasma. Systemic fat embolism is a cause of neurological and general deterioration post-trauma, and may be unrecognized at postmortem examination unless specifically sought.

Pulmonary thromboembolism is a late consequence of trauma especially where relative immobility, vascular injury and the relative hypercoagulable state seen in trauma patients results in the formation of peripheral thrombus.

Systemic Inflammatory Response Syndrome

In 1992, the term systemic inflammatory response syndrome (SIRS) was coined by the American College of Chest Physicians and Society of Critical Care Medicine Consensus Conference as the clinical expression of endogenous mediators of inflammation not necessarily related to infection. The diagnosis of SIRS was defined as the presence of two or more of the following conditions: temperature $>38°C$ or $<36°C$; heart rate >90 beats min^{-1}; respiratory rate >20 breaths min^{-1} or $PaCO_2 < 32$ torr (<4.3 kPa); white blood cells $>12\,000$ cells mm^{-3}, <4000 cells mm^{-3} or $>10\%$ immature (band) forms.

SIRS can be seen in many conditions including sepsis, trauma, pancreatitis, burns or following major surgery.

Severe trauma induces the SIRS as a consequence of (a) hemorrhage with ischemia and subsequent reperfusion (I/R syndrome) and (b) bony fractures or organ damage initiating inflammation and the induction of tissue repair. Inflammatory mediators generated from ischemia/reperfusion injury include proteolytic and oxygen radical metabolites and adhesion molecules, whereas tissue and vessel injury directly stimulate the complement, coagulation and kinin pathways.

In minor trauma the effects of inflammation clearly have a beneficial role in homeostasis and tissue repair. Local tissue injury activates complement and induces tissue macrophages, monocytes and other reactive cell elements to produce various mediators. For example, tumor necrosis factor alpha (TNFα) and interleukin-1 (IL-1) are secreted in large amounts following injury and appear within the circulation within 1 h. These mediators both have significant local and systemic effects. These polypeptide cytokines cause a variety of responses including activation of numerous cell populations and release of secondary cytokines and growth factors. Neutrophil polymorphs and endothelial cells are stimulated to change their activation status and receptor expression leading to adhesion of neutrophils prior to migration and degranulation.

Other factors involved in the early stages of acute inflammation include highly active lipid mediators such as platelet-activating factor (PAF), prostaglandins and leukotrienes which also trigger and perpetuate the local accumulation of inflammatory cells. Complement components (C3a, C5a) are activated to stimulate leukocytes and macrophages to, in turn, increase vascular permeability. These and other components of the inflammatory response are crucial in the repair of injured and infected tissues and should be viewed as a normal physiological response.

It is also important to appreciate that the pro-inflammatory state of the acute phase response also initiates anti-inflammatory mediators. Once the underlying cause of inflammation is controlled and contained by host defenses, often with the assistance of medical and surgical intervention, the pro-inflammatory response will usually subside. For example, macrophages stimulated by various inflammatory mediators will sequentially release various growth factors including fibroblast growth factor (FGF), epidermal growth factor (EGF) and platelet-derived growth factor (PDGF) to induce the formation of newly formed blood vessels and extracellular matrix to begin the process of healing. The production of cortisone and other ‘inflammation antagonists’ are stimulated to ensure the acute phase response is attenuated.

However, in some victims of severe trauma, a massive cellular response is induced and the effects of the inflammatory mediators are not restricted to the sites of direct injury. Thus potent mediators released by various cells including activated macrophages reach the systemic circulation and influence the immune defense, microcirculation and metabolism of remote organ systems. The effect of inflammatory mediators within the systemic circulation underlies the clinical expression of SIRS.

A clinical progression to altered organ function in an acutely ill patient with SIRS such that homeostasis cannot be maintained without intervention defines the MODS. In some patients with MODS organ function cannot be adequately supported and the patient develops MOSF.

The pathologist who performs a postmortem examination on a victim of trauma who succumbed in hospital some days after the initial incident will document various soft tissue and bony injuries. There may be evidence of gastrointestinal hemorrhage from gastric erosions and basal atelectasis and bronchopneumonia within the lungs. It is, however, important for the pathologist to be aware that a complex milieu of various inflammatory mediators had existed within the deceased and which had profound local, cellular and systemic effects on the body.

Adult Respiratory Distress Syndrome

The adult respiratory distress syndrome (ARDS) is clinically characterized by progressive hypoxemia refractory to oxygen therapy and may complicate a variety of medical and surgical conditions and is not necessarily associated with a direct lung injury. Radiographs of the chest reveal bilateral pulmonary infiltrates reflecting protein-rich edema fluid which is caused by an increase in permeability across the endothelial and epithelial barriers of the lung.

The fact that ARDS can occur in cases where there is no evidence of direct lung injury suggests that the systemic effects of inflammatory mediators are involved in the pathogenesis of the disease. A recent review has indicated that inflammatory cells, cytokines and adhesion molecules are important in the initial stages of the syndrome. It is further suggested that ARDS may represent the failure of anti-inflammatory processes within the lung and body as a whole, and the balance of pro- and anti-inflammatory mediators in SIRS may well be the most important issue.

The earliest morphologic change within the lung that may be observed by the pathologist is nonspecific edema. This may be followed by inflammation with necrosis of epithelial cells, fibrin deposition and the formation of hyaline membranes lining alveolar

walls. In the later proliferative phase hyperplasia of type 2 pneumocytes lining the alveolar spaces may be observed. The corresponding macroscopic finding is of heavy firm beefy lung tissue.

In the late or organizing stage, varied amounts of collagen tissue are seen within the interstitium and within alveolar spaces.

Disseminated Intravascular Coagulation

The blood coagulation system protects against loss of blood and is an immediate response to vessel injury. Inside the blood vessel a group of antiplatelet, anticoagulant and fibrinolytic factors restrict activation of the hemostatic system to the site of endothelial injury and thus ensures patency of the vessel. It is in circumstances where this autoregulation becomes insufficient that DIC may develop. DIC has varied etiologies including trauma and is an important factor in the development of MODS.

DIC is characterized by activation of the coagulation system through excessive thrombin formation leading to microthrombi throughout the microcirculation of the body. Consequently there is consumption of platelets, fibrin and coagulation factors and secondary activation of fibrinolytic mechanisms leading to a bleeding diathesis. Therefore DIC may present with signs and symptoms of tissue hypoxia and infarction or as a hemorrhagic disorder subsequent to consumption of the various elements involved in coagulation.

Coagulation may be initiated by either the extrinsic or intrinsic pathway. The extrinsic pathway is triggered by the release of tissue factor or 'thromboplastin'. The intrinsic pathway is activated by contact of factor XII with collagen or other negatively charged surfaces. Both the intrinsic and extrinsic pathways pass through a series of steps which result in the generation of thrombin which in turn converts fibrinogen into fibrin. Thrombin is also a potent activator of platelets and, through the thrombin receptor, thrombin also has direct effects on cell proliferation and regulation of the inflammatory process through expression of both leukocyte adhesion molecules and counter-receptors on endothelium. A recent study has indicated that persistent thrombin activity is closely linked to sustained SIRS in post-trauma patients.

The consequences of DIC are twofold. First, the widespread deposition of fibrin within the microcirculation leads to fragmentation of erythrocytes and obstruction to the microvasculature with resultant ischemia. When patency of the vessel is restored it is to be expected that ischemic damage may lead to inflammation. Secondly, the hemorrhagic diathesis resulting from consumption of platelets and clotting factors, with activation of plasminogen may be the prominent clinical manifestation of the disorder.

Pulmonary and Systemic Fat Embolism

Fat embolism may be an important sequel to soft tissue or bony trauma. Whereas in the past fat globules were believed to be extruded from the edges of bony fractures into the systemic circulation, it is now believed that fat embolism may reflect the instability of lipoproteins in the plasma with coalescence into macroglobules of lipid. Fat globules are not uncommonly seen in lung tissue following fatal trauma or in deaths following orthopedic procedures which are not directly related to the surgery. Pulmonary fat embolism detected microscopically is usually not associated with respiratory failure and the degree of fat embolism must be interpreted in the context of other factors such as the presence of other injuries and underlying natural disease processes.

Globules of fat within the pulmonary circulation may result in vascular obstruction, local vasoconstriction and pulmonary edema. To establish fat embolism as a significant contributing factor to the cause of death one must identify systemic embolism.

It was previously universally accepted that systemic fat embolism occurs when the pulmonary capillaries and veins become 'saturated' with fat globules and thus allow fat to appear within the arterial system. Another possible route for systemic fat embolism is a patent foramen ovale. More recently the suggestion has been made that fat embolism may also relate to an alteration in the activity of lipase or phospholipase which is caused by the embolic fat. This alteration then leads to the precipitation of serum fat in the vessels of the different organs.

Systemic fat embolism results in vascular obstruction and inflammation with petechiae seen within the brain and skin and demonstrable fat on microscopic examination of the brain and kidney using special stains and processing techniques.

Pulmonary Embolism

Individuals who have sustained significant trauma are at risk of peripheral thrombosis and subsequent pulmonary thromboembolism because (a) tissue trauma increases the coagulability of the blood for several weeks, the peak being between one and two weeks; (b) injury to the tissues, especially the legs or pelvic region, may cause local venous thrombosis in contused

muscles or around fractured bones; (c) the injury may confine the victim to bed, either because of general shock and debility, or because the trauma itself necessitates recumbency as in head injuries, severe generalized trauma or injury affecting the legs. Recumbency leads to pressure on the calves and immobility which in turn reduces venous return and stasis because of less muscle massage of the leg veins. These factors combined increase the likelihood of peripheral thrombosis. Thrombus may dislodge from the peripheral circulation to the pulmonary vessels. Depending on the size of the thromboembolus it may impact within the distal pulmonary arteries. In an otherwise healthy individual there may not be any significant consequences. In those with lung or heart disease the affected lung tissue may undergo infarction.

Larger pulmonary thromboemboli may lead to chest pain, shortness of breath, cardiovascular collapse and sudden death. Microscopic examination of the embolic material will allow the pathologist to estimate the age of the thrombus.

Air Embolism

Although air embolism is a relatively rare clinical problem it is of considerable forensic importance because of its association with medical intervention, barotrauma, and occasional association with criminal activity. Air or gas embolism may occur on the venous or arterial side of the circulation and occasionally may occur within both. Venous air embolism is by far the more common clinical and pathological entity.

Lethal air embolism may occur with the introduction of 70–130 ml of air and the rate of administration is said to be important in relation to the clinical consequences. Air becomes trapped within the right atrium and right ventricle resulting in mechanical obstruction to blood flow through the heart because of the 'air lock' produced. In addition, the presence of air or gas within the circulation leads to platelet activation and release of vasoactive substances.

Air may gain access to the arterial from the venous side of the circulation via a patent foramen ovale or through intrapulmonary shunts causing potentially catastrophic obstruction to coronary and cerebral blood flow. However, recent reports have suggested that arterial air embolism may be directly associated with mechanical ventilation as a result of pulmonary barotrauma.

Iatrogenic lethal air embolism can be caused by endoscopic surgery with insufflation of viscera, complications of intravenous catheter insertion and infusion, in surgical procedures of the head and neck and during pulmonary needle biopsy procedures. It follows that other injuries to the head and neck that cause venous injury may also potentially lead to air embolism. Hyperbaric injury or 'the bends' is due to dissolution of nitrogen bubbles within various tissues including blood vessels, as a consequence of inadequate decompression.

The postmortem demonstration of air embolism is difficult. Chest radiographs are essential and recent work has recommended collection of the gas by an aspirometer with subsequent analysis by gas chromatography.

Amniotic Fluid Embolism

Amniotic fluid embolism is a rare and frequently lethal complication of pregnancy characterized clinically by profound and unresponsive hypotension with tachypnea, cyanosis and abnormalities in cardiac rhythm followed closely by the onset of DIC.

Patients may develop amniotic fluid embolism in clinical circumstances ranging from the performance of a second trimester termination of pregnancy up to instances that occur relatively late following completion of labor. The pathophysiology is not completely understood and various mechanisms have been suggested including a mechanical obstruction to pulmonary capillaries, activation of vasoactive substances, the presence and action of tissue factor within amniotic fluid, and anaphylaxis.

The diagnosis is made by the demonstration of squamous cells, mucin, meconium, and lanugo in the maternal pulmonary vasculature. Although the diagnosis has been suggested clinically from aspiration and examination of blood from pulmonary catheters, the diagnosis is generally reached after postmortem examination. Diagnostic material may be seen on conventional hematoxylin-and-eosin stained sections but is better appreciated using special stains such as the Attwood's stain. In addition, immunohistochemical techniques are available to detect squamous cells and more importantly, mucin derived from amniotic fluid.

Tissue and Foreign Body Embolism

Embolization of foreign material within the vascular system is an uncommon but important complication of interventional medical practice. In forensic pathology embolization of projectiles within the vascular system and within hollow viscera is well documented in the literature. In addition, rare cases of hypodermic needle embolization are seen in intravenous drug abusers.

Tissue embolism is well recognized both macroscopically and, more commonly, on microscopic examination in the postmortem examination of individuals dying of severe blunt trauma. Cases in clinical practice of liver tissue embolism to the pulmonary circulation following blunt abdominal trauma have been described. In addition, embolism of cerebral tissue to a lung later harvested for transplantation resulted in clinical problems of coagulation disorder and hemorrhage is reported in the literature.

Cases abound in the medical and forensic literature with regard to various types of embolism from invasive procedures. Cases include fragmented pacing wires, intravascular stents and catheters, dislodgement of orthopedic wires with embolism to the heart and fracture and embolization of mechanical heart valve prostheses. The foreign material may resulting in obstruction to blood flow or perforation of tissue and may act as a nidus for thrombus formation.

In addition to the aforementioned complications of foreign material embolism, surgical procedures such as total hip arthroplasty may result in the embolization of cement (methyl methacrylate monomer), fat and bone marrow debris leading to microthrombus formation and pulmonary vasoconstriction culminating in cardiovascular compromise and occasional death.

See also: **Causes of Death**: Blunt Injury; Sharp Injury.

Further Reading

Bajanowski T, West A and Brinkmann B (1998) Proof of fatal air embolism. *International Journal of Legal Medicine* **111**:208–211.

Gando S, Kameue T, Nanzaki S, Hayakawa T and Nakanashi Y (1997) Participation of tissue factor and thrombin in post-traumatic systemic inflammatory syndrome. *Critical Care Medicine* **25**:1820–1826.

Hasleton PS and Roberts TE (1999) Adult respiratory distress syndrome. *Histopathology* **34**:285–294.

Herron DM, Vernon JK, Gryska PV and Reines HD (1999) Venous gas embolism during endoscopy. *Surgical Endoscopy* **13**:276–279.

Hidekazo O, Kobayashi H, Hirashami Y, Yamazaki T, Kobayashi T and Terao T (1998) Serological and immunohistochemical diagnosis of amniotic fluid embolism. *Seminars in Thrombosis and Haemostasis* **24**:479–484.

Ibrahim AE, Stanford PL and Freund PR (1999) Pneumothorax and systemic air embolism during positive-pressure ventilation. *Anaesthiology* **90**:1478–1481.

Rose S and Marzi I (1998) Mediators in poly-trauma. Pathophysiologic significance and clinical relevance. *Langenbeck's Archives of Surgery* **303**:199–208.

Urban MK, Sheppard R, Gordon MA and Urquhart BL (1996) Right ventricular function during revision total hip arthoplasty. *Anesthesia and Analgesia* **82**:1225–1229.

Traffic Deaths

M J Dodd, Victorian Institute of Forensic Medicine, Southbank, Victoria, Australia

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0432

Introduction

The investigation of motor vehicle accidents resulting in fatality and pedestrian deaths constitutes a significant percentage of the work load of a busy forensic facility.

A firm understanding of the pattern of injury in these cases is a fundamental prerequisite and is heavily relied upon in the law courts, as frequently, the pathologist may be called to give evidence, or at the very least, to furnish a report which adequately addresses the questions posed by the investigating officers.

The purpose of the forensic pathologist in investigation of road and traffic fatality is as follows:

1. to establish the cause of death;
2. to determine the pattern of injuries;
3. to note the presence of any natural disease;
4. to determine the role (if any) of alcohol or drugs;
5. to report these conclusions based on a full examination.

In many cases the cause of death is self-evident. Frequently, after high speed collision, the body incurs extensive trauma and disruption. In many cases, however, the external examination fails to disclose an adequate cause of death, or at the very least, provides evidence of minimal trauma. It is fundamental that a definitive cause of death be provided and it is only after the completion of an adequate autopsy that this requirement is fulfilled. The determination of the pattern of injuries is crucial in the reconstruction of the series of events leading to the fatality. This information is relied upon heavily by trained investigating officers and may be crucial at the time of a coroner's investigation or possibly the hearing of cases in the higher courts.

The presence of significant naturally occurring disease must also be disclosed, as a small number of cases of road trauma resulting in death may be directly attributed to an overwhelming medical event. In these cases, the event is frequently witnessed, either by a passenger or by a pedestrian or driver nearby. Frequently, the vehicle is seen to waver or lose control and not infrequently, the driver is able to pull the car over to a safe position at the side of the road prior to collapse or death. In these cases, the cause of death is

not infrequently one of cardiovascular origin, but in other cases, the disease process may be more obscure.

Of great significance in the investigation of road fatality is the role of alcohol, on its own, or in combination with drugs, prescription or otherwise. Of increasing significance today is the influence of illicit drugs such as the opioids and cannabinoids.

The final written report must include an adequate list of injuries and the influence of these injuries on the organ systems, comment regarding the presence of natural disease and the influence of alcohol or drugs, if any. It is this final report that is used by the coroner or medical examiner and its worth is dependent on the skill of the pathologist in charge of the case.

The ideal situation would dictate that the forensic pathologist attend every road fatality. Clearly this is not feasible given the number of fatalities per annum and the number of pathologists employed in a given metropolitan institution. Interpretation of injuries, however, is often facilitated by a personal visit to the scene or inspection of the vehicle or vehicles involved. Failing this, the inspection of the scene photographs is often invaluable.

The pattern and severity of injuries are essentially dictated by:

1. the direction of impact;
2. the design of the cabin;
3. the force of the impact;
4. the behavior of the vehicle after the impact;
5. whether the driver/passenger was injected from the vehicle;
6. whether a restraining system was in use (seat belt or airbag);
7. whether death may have been compounded by a factor such as fire or inhalation of toxic fumes.

The injury pattern may, for simplicity, be divided into those seen in drivers and passengers.

Driver Injury

In the event of a high speed head-on collision, in addition to rapid deceleration, the car frequently dips, causing the driver to slide forward into the steering column and at the same time, elevate, allowing the head to impact with the fascia. This complex movement results in flexion of the cervical and thoracic spines and at the moment of contact with the head to the fascia, lead to compression of the axial spine. The axial compression may be such that the pelvic girdle may fracture. The driver, frequently being aware of the impending situation, may adopt a 'bracing posture'. The bracing posture is typified by outstretched forearms and a gripping action on the steering wheel, together with a forced rigidity and straightening of the braking leg.

As seen with axial spinal compression, the force may be transmitted from the braking foot, through ankle, leg and ultimately into the acetabulum and pelvis. This pattern of fracture is frequently invaluable in determining driver from passenger when both parties are ejected from the vehicle and particularly, when restraining devices are not used.

Variations on the classic injury pattern may be caused by intrusion of various vehicle parts into the cabin such as the engine, steering column and firewall.

In the situation of rear impact at speed, the cervical spine may undergo severe hyperextension and then later, deceleration with hyperflexion. This complex range of movements may cause substantial trauma to the spinal cord and supporting axial spine and musculature.

Side impacts may show any combination of the above.

Who Was the Driver?

To discriminate driver from passenger in a scenario where restraining devices have been employed, a dedicated external examination is required to disclose the typical distribution of bruising and abrasion caused by lap/sash belts. In many cases, the bruising is immediately apparent, but in others, subcutaneous dissection may be required to disclose points of maximal contact and pressure (**Fig. 1**).

As suggested earlier, frequently the driver is aware of the impending situation whereas the passenger is often oblivious. The presence of bracing type leg fracture and/or pelvic fracture may adequately discriminate between driver and passenger. Indeed, the

Figure 1 (see color plate 9) Typical lap/sash passenger distribution of bruising and abrasion.

bracing actions may well be demonstrated in the driver at the level of forearm, wrist and shoulder whereas this pattern may be entirely absent in passengers.

A dedicated examination of the deceased (and clothing) by a crime scene investigator may be rewarding. The discovery of glass, automotive paint, plastic and fabric may help to accurately place the deceased in the vehicle, and therefore assist in discerning driver and passenger.

Pedestrian Death

Pedestrian fatalities may account for more than 50% of all road-related deaths. Many texts have classified the pattern of injuries seen in pedestrians into primary and secondary. The primary injury is defined as the injury caused by the first impact of the vehicle on the victim whereas the secondary injury is caused at the time of impact with the ground. Other texts have referred to primary, secondary, tertiary and quaternary injuries. This rather complex classification may suggest to coroner or barrister that these injuries are easily discriminated and therefore, at the time of the inquest, may be frankly misleading. This classification discriminates the injuries thus as follows.

1. The primary injury is caused by initial impact with the vehicle.
2. The secondary injury is caused by the initial impact with the vehicle once the body has been propelled on to it (i.e. onto the bonnet or windscreen).
3. The tertiary trauma is caused by contact with the ground after the body has left the vehicle.
4. The quaternary injury is caused by further vehicles traveling in either the same or opposite directions.

It is often not fully appreciated that in many cases, after impact with a conventional vehicle, the victim is 'scooped' up onto the bonnet and then despatched to the ground either across the bonnet or over the roof towards the rear of the vehicle. This scenario is especially pertinent in cases where application of the brakes by the vehicle immediately prior to collision has caused the front end to dip somewhat.

Naturally, larger vehicles such as vans and trucks tend to impact on the victim in a full frontal fashion leading to a forward propulsion. The victim may then go under the vehicle and be crushed by the wheels or damaged by protruding parts of the vehicle. This rather simplistic outline needs to be tempered with the consideration of the height of the victim (i.e. adult or child).

Leg trauma is identified in approximately 85% of pedestrian fatalities. Indeed, 25% of fatal outcomes disclose evidence of compound fracture to tibia and/or fibula. The classic texts frequently describe fracture patterns which are purported to indicate the activities of the victim immediately prior to impact. It is said that if the fracture is oblique in orientation, it indicates weight bearing at the time of impact. A transverse fracture is said to indicate a nonweight-bearing state. By inference, if the points of fracture are at different levels it would suggest that the victim was either walking or running prior to impact. The point of a fracture wedge (often seen on a radiograph) is said to indicate the direction of force.

In reality, the fracture patterns may be complex and therefore do not readily lend themselves to such a simplistic approach. It is suggested that overinterpretation of fracture patterns may, in many cases, be misleading and therefore counterproductive. Additional autopsy techniques are often required. Measurements of the distance from heel to level of abrasion/laceration and fracture line may assist in the estimation of the bumper bar height. Subcutaneous dissection of the lower extremities frequently discloses a zone of hemorrhage at the point of impact, and this can readily be measured from the heel. Subcutaneous dissection of the back may also be required.

Motor Cycle Fatality

The motor cycle fatality may constitute approximately 10% of all road casualties. A review of motor cycle-related injuries in children and adolescents, funded by the Victorian Health Promotion Foundation and Accident Research Centre cited ten factors implicated as contributing to fatality.

1. Being male
2. Being young
3. Being inexperienced
4. Being unlicensed
5. Riding on a borrowed motor cycle
6. Consumption of alcohol
7. Riding in peak hour
8. Negotiation of curves
9. Slippery or uneven surfaces
10. Poor motor-cycle maintenance

The majority of motor cycle accidents appear to occur in the daylight hours (65%). The engine capacity was deemed to be a significant factor with 54.6% of fatalities involving motor cycles with an engine displacement of 250–500 cc.

As would be expected, trauma to the head, chest and abdomen is commonplace. However, a frequently overlooked and unsuspected facet of injury in motor cycle riders is one of cardiac trauma. One study reviewed 546 autopsies in patients who had incurred nonpenetrating traumatic cardiac injury. All of these patients were deemed to have died from their

cardiac or associated injuries and of the 546 cases, 64.6% died from cardiac rupture. Of these, 30% demonstrated multiple chamber rupture and the remainder were associated with rupture of the aorta.

Patients incurring such injury may present with alterations of the ECG, myocardial failure, cardiac tamponade or hemothorax. These injuries may present without obvious rib fracture or chest wall deformity and therefore may elude primary diagnosis.

Rupture of the right atrial appendage is not uncommonly demonstrated in blunt trauma to the chest and abdomen. (**Fig. 2**). The underlying cause for the traumatic disruption of the thin-walled right atrial appendage is deemed to be a 'water hammer' like column of blood, the pressure being generated by blunt compression of the inferior vena cava and viscera leading to sudden and catastrophic distention, and ultimately, disruption of the appendage.

Trauma to the head and neck is a frequent and significant injury type, constituting over 50% of all road traffic accidents. The trauma ranges from relatively minor, in the form of laceration or abrasion, to simple or complicated forms of skull fracture. The hinge-type base of skull fracture is commonly seen as a sequel to blunt lateral head impact (**Fig. 3**).

The cervical spine may be traumatized leading to dislocation of the atlanto-occipital joint, brainstem injury, frank fracture, pontomedullary rent or disruption of the vertebral arteries. The latter may lead to significant basal subarachnoid hemorrhage.

Thoracic trauma may constitute frank cardiac rupture, cardiac contusion, hemopericardium, traumatic rupture of the aorta (particularly of the distal arch segment), hemopneumothoraces, rib fracture (frequently a flail segment), pneumothorax, pulmonary laceration and contusion and disruption of the hilum.

Figure 2 Rupture of the right atrial appendage.

Figure 3 (see color plate 10) Hinge-type base of skull fracture commonly seen as a sequel to blunt lateral head impact.

Abdominal injuries, in order of frequency, constitute laceration or contusion of liver, spleen, kidney, mesenteries, pancreas, gut and diaphragm. A significant number of cases may manifest as delayed rupture of the spleen, leading to catastrophic hemorrhage days or possibly weeks after the event.

Pelvic injury may be incurred as a result of direct trauma to the pelvis either by anterior impact or transmitted force from fractured femora. A fractured pelvis may lead to rupture of the urinary bladder and intense hemorrhage as a result of torn iliac vessels. This particular form of trauma carries a significant morbidity and mortality.

Complicating Factors

A small but significant number of fatalities may be directly related to causes other than immediate physical trauma at the time of impact; for example inhalation of carbon monoxide or cyanide which may be liberated from combustion of synthetic materials in the cabin. Traumatic asphyxia from compression of the chest by steering column or overturned vehicle is occasionally encountered.

In many cases, the ultimate cause of death may be thermal injury. Evidence of sustained respiration in these cases may be demonstrated by the presence of black particulate material within the major airways and further verified by analysis of carboxyhemoglobin in postmortem blood specimens. In cases of explosion or flash fire, such findings may be absent.

In many cases, an autopsy is requested on a person who has sustained trauma but survived for a time in

hospital or convalescent home. These often constitute common entities such as pneumonia, or the effects of fat embolism syndrome. Rare cases may include cardiac fibrosis and arrhythmia secondary to healed cardiac contusion.

Pre-existing Disease

The rare and problematic cases constitute those of pre-existing natural disease. These cases account for approximately 1% of fatalities overall and are often typified by death behind the wheel with minimal apparent external evidence of trauma. The event is often witnessed and the collision, if any, is at low speed. These fatalities may disproportionately represent the older population and in many cases, significant medical and surgical disease is well documented. Direct communication with the treating medical practitioner is often very rewarding.

Such conditions may include acute myocardial infarction in a person with known significant coronary artery disease, a cerebrovascular accident, transient ischemic attack or more uncommonly, diabetes or epilepsy. The diagnosis of epilepsy must be based on a confirmed antemortem diagnosis unless a definitive anatomical lesion can be demonstrated either macroscopically or histologically. Epilepsy may be suspected if the deceased shows evidence of convulsion, such as biting of the tongue and incontinence. In confirmed epileptics, the presence of anticonvulsant medications must be sought and quantified.

The worth and necessity of toxicological analysis of body fluids in cases of road fatality cannot be over emphasized. All too frequently, the influences of drugs and alcohol are manifest on the roads each year. A worrying trend in road fatality now appears to be the usage of prescription medications, such as the benzodiazepine group which act in a synergistic fashion with alcohol to reduce the ability to adequately control a motor vehicle. The modern laboratory requires sampling of not only blood but bile, urine, liver, gastric content and vitreous humor for such analyses. The sampling of blood from peripheral veins is preferred as blood taken from body cavities may be contaminated by the diffusion of drugs from the gastrointestinal tract. Cardiac and cavity blood may be used for the analysis of carbon monoxide if peripheral blood cannot be obtained.

Investigation of the 'Hit–Run' Fatality

The hit-and-run fatality is an uncommon but highly significant event requiring ultimately a multiskilled approach to the examination, both at the scene and in the postmortem room. The conduct of the examination for the investigation of the hit-and-run fatality epitomizes the concept of the team effort.

The body should be received in the mortuary fully clothed and strict instructions should be given to the mortuary attendant that the clothes should remain *in situ* prior to examination by the pathologist. A full photographic profile needs to be undertaken to document the state of the clothing and in particular, trace evidence such as plastic, glass and paint. These telltale exhibits must be collected, protected and handed over to the investigating authorities with a reliable and demonstrable chain of custody protocol. The clothes may be removed and retained as evidence.

A full photographic profile of incurred injuries should be documented; particularly the use of close-up shots, with and without scale, is required. Particular note should be taken of patterned injuries such as bruises and abrasions which may provide an indication of make and model of the suspected vehicle.

The full autopsy is performed with due regard to accurate description of all injuries in relation to anatomical landmarks. The location, type, dimension and depth of all injuries need to be meticulously recorded. The autopsy may also require subcutaneous dissection of the legs to indicate accurately the level of contact with bumper bars or other protruding parts of the offending vehicle (**Figs 4** and **5**). Radiology may be required to document fractures not identified by open dissection. It is recommended that foreseeable problematic cases have at the very least, a radiological survey of the lower limbs. The opinion of a specialist radiologist may be required.

Figure 4 (see color plate 11) Bruising and abrasion indicating the level of contact with bumper bars.

Figure 5 (see color plate 12) Subcutaneous dissection of the legs to determine accurately the level of impact.

As indicated earlier, a full toxicological analysis of body fluids is required to demonstrate the influence of alcohol or drugs. In some cases, the absence of a particular medication may be as relevant as the presence of a drug in excess.

Conclusion

In closure, it is stressed that the forensic pathologist should impart his or her knowledge and experience to those involved in the investigation of motor vehicle and pedestrian death such as specialist police accident investigation officers and coroner's staff. The collective experience of a forensic institute is also an invaluable resource in the teaching of undergraduate medical students and those in the legal profession.

See also: **Accident Investigation**: Driver Versus Passenger in Motor Vehicle Collisions; Motor Vehicle. **Alcohol**: Blood; Body Fluids; Interpretation. **Anthropology**: Skeletal Trauma.

Further Reading

Cumberland G, Riddick L and McConnell C (1991) Intimal tears of the right atrium of the heart due to blunt force injuries of the abdomen. *American Journal of Forensic Medicine and Pathology* **12**:102–104.

Haworth N, Ozane-Smith J, Fox B and Brunnen I (1994) *Motor-cycle Related Injuries to Children and Adolescents*. Monash University Accident Research Centre Report no. 56.

Mant AK (1993) Injuries and death in traffic accidents. In: Mason JK (ed.) *The Pathology of Trauma* 2nd edn, pp. 1–16. London: Edward Arnold.

Noon G, Boulafendis D and Beall A (1971) Rupture of the heart secondary to blunt trauma. *Journal of Trauma* **11**:122–128.

Parmley L, Manion N and Mattingly T (1958) Nonpenetrating traumatic injury to the heart. *Circulation* **18**:371–396.

Certification *see* **Accreditation of Forensic Science Laboratories**.

Chain of Evidence *see* **Crime-scene Investigation and Examination**: Collection and Chain of Evidence.

CHEILOSCOPY

J Kasprzak, Military Forensic Laboratory, Warszawa, Poland

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0436

Definition

Cheiloscopy is a forensic investigation technique that deals with identification of humans based on lips traces.

History

The biological phenomenon of systems of furrows on the red part of human lips was first noted by anthropologists. R. Fischer was the first to describe it in 1902. Until 1950, however, anthropology merely mentioned the existence of the furrows without suggesting a practical use for the phenomenon. Since 1950 the Japanese have carried out extensive research in the matter. In the period 1968–1971 two Japanese scientists, Y. Tsuchihashi and T. Suzuki examined

1364 persons at the Department of Forensic Odontology at Tokyo University. Based upon that research it was established that the arrangement of lines on the red part of human lips is individual and unique for each human being. This statement led to the conclusion that there is the possibility of using the arrangement of furrows (on a trace, in a linear form) on lips for the identification of a person. In further research the Japanese scientists examined the principles of the heredity of furrows on the red part of lips. It was in Hungary during 1961 that the first research in Europe was carried out in the subject of lip prints. The examination started after lip traces had been found on a glass door at the scene of a murder. It was established later that the murderer pushed the door open with his head because his hands were covered in blood, leaving a trace of the red part of his lips on the glass. In the process of examination of the lip traces their usefulness for criminalistic identification was proven.

Cheiloscopic research was also carried out by specialists in anthropology, odontology, forensic medicine and forensic science, in Brasil, Iran, Hungary, France, Germany, Italy, United Kingdom, the Soviet Union and Czechoslovakia. The results of this research constitute the proof of lip-print individuality and also of its usefulness for criminalistic identification.

In Poland, the interest in lip prints started in 1966 when a lip print was revealed on window glass at the scene of a burglary. Research was carried out, and its results were comparable to those achieved in Japan and Hungary.

The research was only of preliminary character and did not allow for practical application of results as yet. There was a need to develop one cohesive cheiloscopy system, practicable in forensic cheiloscopy.

A project aiming at that objective was launched in 1982, in the Forensic Institute of Warsaw University Criminal Law Department, in cooperation with the former Forensic Institute of Militia in Warsaw. The material for study was collected in the former Military Training Center at Mińsk Mazowiecki. Lip prints were collected from 1500 persons (including 107 women), coming from various locations around the country. The age of the volunteers varied from 5 to 60 years. Altogether more than 7000 traces of the red part of the lips were examined. As a result of the examination the individuality of lines in the red part of lips and their unchangeability within the limits practicable for identification was proven. The examination determined methods for revealing and securing the traces of the lips, methods of acquiring the comparative material, and, more importantly, detailed methods of cheiloscopic expertise. The possibilities of registration of traces of the red part of lips were also examined, and a file comprising 1500 individuals was organized on a laboratory scale.

Since 1985, in Poland, the methods of finding and recovery of lip traces, recovering comparative material, and the techniques employed to carry out that expertise have been introduced into casework of the Fingerprint Department, of the Central Forensic Laboratory of Police in Warsaw. During the years 1985–1997, cheiloscopic techniques have been used in 85 cases, including 65 burglary cases, 15 cases of homicide, and five cases of assault. In 34 cases the identification was positive, which means that cheiloscopic techniques were equal in value to other types of forensic evidence. It has also been included in evidence for presentation in court.

The Subject and Scope of Cheiloscopy

It is not easy to place the lip prints in the general system of traces. A lip print may be revealed as a stratified surface trace with visible elements of lines (furrows). In this case a trace has the shape of a print – the individuality of its properties make it possible to individually identify a human being. In the case when the lines are not clear (only the shape of the lips is printed), individual identification of a human being based on this trace is extremely difficult (unless the trace contains more individual characteristics, e.g. scars), and often identification ends with group identification. In these cases it is possible to examine the substance which constitutes the trace e.g. saliva, as a biological trace. The potential for DNA typing from the lip print is evidentially here. However, this process has so far not been attempted. If the lips are covered with remains of food or cosmetics and the lines are printed unclearly, the trace will have the shape of a stain, and can be subjected to chemical examination, in order to determine the substance covering the lips. The value of such a trace will depend on its type. Traces with clear lines and individual elements enable individual identification of a human being. In this sense lip prints have the same value as dactyloscopic (fingerprint) traces. In the case of traces in the shape of stains the identification examination terminates with group identification, in their character they are similar to other chemical and biological traces (including DNA identification).

Apart from identifying and evidential use, lip prints may also be used in detection work, being the source of tactical and criminalistic information. A lip print at the scene of a crime can be a basis for conclusions as to the character of the event, the number of the people involved, sexes, cosmetics used, habits, occupational traits, and the pathological changes of lips themselves.

The Use of Lip Prints in Crime Detection

A trace of this kind can only be revealed at the point of direct, physical contact of the perpetrator's lips with an object at the scene of a crime. This trace should be looked for on cutlery and crockery items, particularly if a meal was eaten at the scene of a crime.

In practice, lip prints have also been revealed on the surface of windows, paintings, doors, plastic bags and cigarette ends.

Finding and recovery of lip traces is not difficult. The techniques used in fingerprinting are very good for the purpose. The easiest method makes use of fingerprint powders and fixing on foil. Aquaprint and cyanoacrylamide may also be applied.

The taking of control prints from the suspect (comparative material) is also not difficult. The method of obtaining the traces was the following. The examined person covered the mouth with a thin layer of skin care cream (e.g. Nivea). After about 3 minutes, a strip of paper 120 mm long and 45 mm wide mounted on a specially profiled roller (made from fingerprinting roller) was lightly pressed to the lips. The impression was subsequently visualized with the use of ferromagnetic powder used in developing latent fingerprints, and then fixed on transparent foil.

In the course of research a classification of patterns of the lines of the red part of lips was made. As the basis for classification, the middle part of the lower lip, 10 mm wide, was taken. This fragment is almost always visible in the trace. The determination of the pattern depends on the numerical superiority of properties of the lines on the fragment (**Fig. 1**). If the lines prevail, the pattern is described as linear, 'L'. If the bifurcation is dominant, it is called bifurcate, 'R'. If the lines cross, the pattern is dubbed reticular, 'S'. In the case when no superiority can be established, the pattern is named undetermined, 'N'. This classification has been proven in practice.

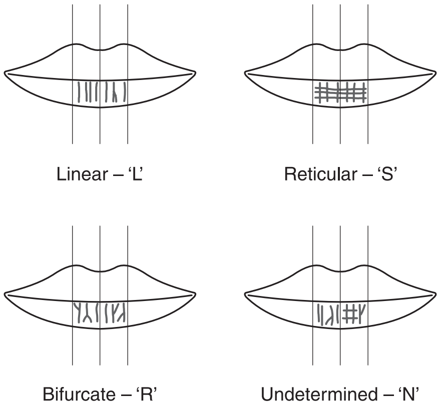


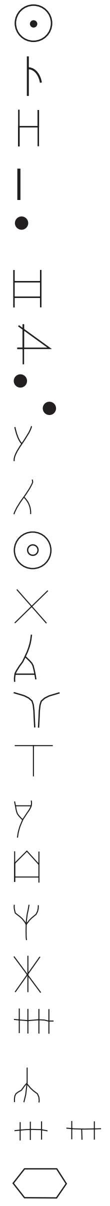
Figure 1 Types of patterns on the red parts of lips.

The next step in the identification is establishing individual features of the pattern of the lines. The discussed study resulted in the first catalogue of 23 types of individual features. The types of features were differentiated as a result of microscopic examination of lip prints, collected from 1500 individuals. The types of features were listed separately (**Table 1**) but in practice the features occur in combinations of two or more. An element of a feature, or the whole feature may appear as a part of another feature,

Table 1 Individual features of line pattern on red part of lips

Type of features
1. An eye
2. A hook
3. A bridge
4. A line
5. A dot
6. A rectangle-like
7. A triangle-like
8. A group of dots
9. A simple top bifurcation
10. A simple bottom bifurcation
11. A double eye
12. Crossing lines
13. A closing bottom bifurcation
14. A delta-like opening
15. A simple opening
16. A closing top bifurcation
17. A pentagonal arrangement
18. A branch-like top bifurcation
19. A star-like bifurcation
20. A fence
21. A branch-like bottom bifurcation
22. A double fence
23. A hexagonal arrangement

Graphic symbol



which is the main difference between properties of system of furrows on lips and those of fingerprint lines (**Fig. 2**).

In order to examine carefully the number of individual features appearing on the trace, their distribution and frequency of occurrence, 100 traces were chosen at random from each kind of pattern. A total of 400 lip prints were analyzed in detail. The examination required the observation of the trace and the count of noticeable features identical with those 23 distinguished individual features in the catalogue. In the 400 lip prints examined 456 215 individual properties were found which gives an average of 1145.5 individual features per trace. Such a high number of individual properties appearing on a lip print opens for cheiloscopy new previously unperceived venues in the scope of personal identification.

The research discussed above has led to the creation of a set of useful practical methods of lip print analysis. There is already a 'protocol' for carrying out a cheiloscopic examination:

1. The method of determining common features;
2. The method of photographic montage;
3. The contour method.

The determining of common properties is the most basic identification method. The essential part of this method is the thesis that in order to establish identity between evidential and comparative material one should determine the common features, including individual features from the catalogue of 23 types of features (Table 1). According to the theory of combinations, using the same principles as in fingerprint analysis, in order to established identity of an evidential trace with a comparative trace, one should determine seven individual features for Poland and nine features for the whole world. However in Polish forensic practice, usually 15–20 common features are distinguished.

The photographic montage method and the contour method supplement and support the method of establishing common properties. In the method of a photographic montage, a fragment of the photograph of a comparative trace is fitted to a photograph of an evidential trace and the conformability of individual properties is examined. In the contour method, contours of the most characteristic arrangements of lines, on the red part of the lips are transferred onto transparent foil and are then compared with the lines in the photograph of evidential and comparative traces.

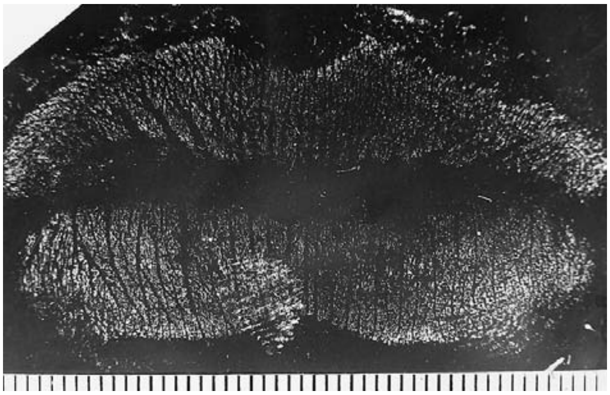


Figure 2 Lip print from the scene of a burglary.

The practical use of lip prints in detection work shows that the trace of this kind carries a huge amount of precious information which can be used in the reconstruction of the event, establishing versions and checking them and identifying suspects.

Full utilization of lip prints depends to a high degree on the skill of members of the law enforcement agencies. The problems involved in cheiloscopy are relatively little known and thus, so far, lip prints have been used only occasionally, despite their frequent occurrence on the scene of the crime.

See also: **Crime-scene Investigation and Examination**: Recording; Recovery of Human Remains. **Deoxyribonucleic Acid**: Future Analytical Techniques. **Fingerprints (Dactyloscopy)**: Identification and Classification.

Further Reading

Afchar-Bayat M (1979) Klassifizierungsystem für Lippenabdrücke. *Kriminalistik* 3(01).

Domachowska R (1979) Z problematyki identyfikacji śladów warg (Problem of identification of the lip prints). *Problemy Kryminalistyki* **137**.

Endris R and Poetsch-Schneider L (1985) Zum Beweiswert des menschlichen Lippen – und Nagelreliefs bei Identifizerungen. *Archiv für Kriminologie* **175**:1–2.

Illàr S (1972) Az Ajaknyomokról. *Belügyi Szemle* **11**.

Kasprzak J (1984) Linie czerwieni wargowej – cechy biologiczne i ich znaczenie w badaniach idenyfikacji człowieka (The lines on the red part of lips – biological properties and their importance for human identification). *Problemy Kryminalistyki* **163**.

Kasprzak J (1988) Methoische Grundfragen der cheiloskopischen Begutachtung. *Kriminalistik und forensische Wissenschaften* **69**:70.

Kasprzak J (1988) Cheiloskopie – eine neue kriminalistische Methode zur Identifizierung von Personen. *Kriminalistik und forensische Wissenschaften* **69**:70.

Kasprzak J (1989) Ujawnianie i zabezpieczanie śladów czerwieni wargowej (The revealing and securing traces of red part of lip). *Zeszyty Metodyczne* **37**.

Kasprzak J (1990) Possibilities of cheiloscopy. *Forensic Science* **46**.

Kasprzak J (1991) *Chejloskopia kryminalistyczna (Forensic Cheiloscopy)*. Warszawa.

Kasprzak J (1991) *Ślady czerwieni wargowej* (Traces of the red part of lip). Legionowo.

Kasprzak J (1993) Wiek człowieka, a niezmienność układu linii czerwieni wargowej (The human age and unchangeability of lines on the red part of lips). *Problemy Kryminalistyki* 200–201.

Kasprzak J (1996) Badania chejloskopijne w polskiej praktyce kryminalistycznej (Cheiloscopic research in Polish forensic practice). *Problemy Współczesnej Kryminalistyki*. Warszawa.

Reiterova S (1986) Otazky cheiloskopie. *Acta Universitatis Palackianee Olomucensis* **IV.**

Tsuchihashi Y (1974) Studies on personal identification by means of lip prints. *Forensic Science* 3(3).

Chemical Testing *see* **Analytical Techniques**: Presumptive Chemical Tests.

Child Abuse *see* **Clinical Forensic Medicine**: Child Abuse.

Chromatography *see* **Analytical Techniques**: Hyphenated Chromatographic-Spectroscopic Techniques; Separation Techniques.

Class *see* **Evidence**: Classification.

CLINICAL FORENSIC MEDICINE

Contents

Overview
Child Abuse
Defense Wounds
Evaluation of Gunshot Wounds
Recognition of Pattern Injuries in Domestic Violence Victims
Self-inflicted Injury
Sexual Assault and Semen Persistence

Overview

S Pollak, University of Freiburg/Br, Freiburg, Germany
P J Saukko, University of Turku, Turku, Finland

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0437

Introduction

In the wider sense, clinical forensic medicine means the application of clinical knowledge and skills to living individuals corresponding to the special needs of the respective legal, judicial and police systems. The organization of clinical forensic work differs considerably from place to place, partly depending on the legal situation (e.g. the so-called 'adversary system' of criminal procedure in the Anglo-Saxon tradition on the one hand and the 'inquisitorial system' in Continental Europe on the other). It is therefore impossible to deal with any regional peculiarities, and this article is necessarily confined to a general view of the main fields of work usually done by forensic medical examiners, sometimes also called

'forensic physicians', 'forensic medical officers', 'police surgeons' or 'clinical forensic practitioners'. **Table 1** shows typical reasons for their attendance although the focal points of demand may vary to a large extent.

From the topics listed in Table 1 it becomes obvious that clinical forensic medicine comprises a large spectrum of duties which are usually handled by doctors of several disciplines involving the medicolegal interface: legal medicine (as a 'stand-alone' subject in the undergraduate curriculum at medical schools of Continental Europe), forensic pathology, psychiatry, emergency medicine, pediatrics, gynecology, public health and others.

Some aspects of clinical forensic medicine are dealt with in dedicated articles and will not be discussed here.

Clinical Examination of Living Victims

Victims of an alleged crime or suspected perpetrators often have to be examined with regard to the presence of injuries. The observations and the medical report on the wounds is likely to play an important part in any subsequent legal proceedings. Therefore, the physical examination and the documentation of the relevant results must be performed in an adequate and accurate manner. Descriptions and conclusions should be phrased in terms which are also intelligible to lay persons. If scientific language is inevitable, an explanation must be given.

It is impossible to prescribe a standardized format to suit every case, as the circumstances differ so much. Apart from the actual examination procedures, it has to be decided if samples need to be taken (for instance blood for alcohol or drugs, genital swabs, urine, hair samples and so on). In most cases photography is helpful and desirable, especially in complex and patterned injuries. Sometimes even the absence of visible injuries might be important, e.g. in false allegations. For detailed recording of small marks such as tiny petechiae a magnifying glass or an operation microscope is necessary. Whenever possible, the whole

Table 1 Categories of clinical forensic work

1. Examination of living victims:
   (a) bodily injury due to criminal assault
   (b) rape or other sexual crimes in adults
   (c) physical and sexual child abuse
   (d) spouse abuse and other kinds of domestic violence
   (e) abuse of elderly persons
   (f) torture
2. Examination of suspected perpetrators
3. Examination of self-inflicted injuries
4. Medical investigation in traffic accidents:
   (a) examination of pedestrians
   (b) determination of driver vs. passenger
5. Examination for fitness to drive:
   (a) assessment of impairment to drive due to alcohol and/or drugs
   (b) specimen taking (blood samples)
6. Assessment of the effect of drink and/or drugs on responsibility
7. Mental health assessment
8. Assessment of fitness to be detained and interrogated
9. Assessment of physical ability required for work
10. Reports, statements, expertise:
    (a) written and photographic documentation of medical findings
    (b) interpretation of wounds and other medical evidence
    - conclusions as to the causation of injuries (e.g. which type of weapon has been used?)
    - assessment of the severity of bodily harm and its dangerousness
    - identification of offensive and defensive injuries
    - medicolegal reconstruction of the circumstances and the course of events
11. Presentation of medical evidence in court:
    (a) witness as to fact
    (b) professional witness
    (c) expert witness
12. Medical care of detainees:
    (a) short-term custody (police)
    (b) long-term custody (prison)
13. Health care of police officers
14. Determination of age

body from head to toe including the genitalia should be minutely inspected in a good light. In order to prevent any reproaches it is recommended that a chaperone should be present whenever a male physician investigates a female patient.

Each medical report should contain some basic information on the victim, i.e. size, stature, body weight etc. All remarkable findings must be recorded and their location defined in relation to easily identifiable marks or fixed body parts (e.g. the middle line and the distance from the sole in upright position). The size and shape of each wound should be detailed by precise measurement using a ruler or a tape measure. If there is a multitude of wounds, it might be advantageous to group them according to their kind and severity or to their location within anatomical regions. Body diagrams or sketches can be helpful.

The margins of each wound require close inspection and accurate description: whether they are lacerated or incised, whether they are shelved or excoriated, whether there is bridging or embedment of foreign material (e.g. soil, glass, paint). In addition, the examiner has to take note of the wound edges (sharp or blunt ends), of concomitant bruises and all other features which may help to distinguish between lacerations, incised wounds and other kinds of penetrating injuries. The condition of the wounds must also be assessed with regard to surgical treatment, infection and signs of repair.

Similarly, other types of external injuries such as abrasions and bruises have to be reported with reference to their location, size, shape and appearance. Sometimes characteristic details may be imprinted. In fresh abrasions the skin tags sometimes indicate the direction of the blunt force. Bruising mostly concerns the subcutaneous layer, often in combination with intradermal extravasations (**Figs 1** and **2**); the latter are patterned if the skin has been pressed and squeezed into grooves (for example by the impact of a rubber sole). Parallel 'tram-line' bruises derive from the impact of a stick or rod, which is of special importance in suspected physical child abuse. Nevertheless, it should be emphasized that most bruises do not have any characteristic shape. Immediately after contusion, the skin need not show any changes of color; in cases of suspected blunt trauma it is recommended to wait one or two days and to examine the victim a second time when extravasation has extended enough to become visible. It is well known that color changes occur within a few days so that the initially bluish-red bruise becomes greenish-yellow.

Figure 1 Bruises on the left thigh of a young woman from multiple blows with a looped extension cord.

Figure 2 Surviving victim with numerous pellet injuries of the left upper arm and breast from a distant twelve-bore shotgun discharge without penetration of the thoracic wall.

Some major points of external victim examination are specified in **Table 2**.

Injuries of Surviving Victims and of Assailants Following Attempted Manual or Ligature Strangulation

One of the common tasks of police surgeons and forensic medical experts is to examine persons who are supposed or claim to have been strangled by hand or by ligature: for instance in sex-related offences such as rape, in attempted homicide and due to maltreatment in cases of domestic violence.

In manual strangulation the pressure on the neck is exerted externally by the hand(s) or the forearm resulting in occlusion of the blood vessels and the air passages of the neck. Most victims exhibit typical skin marks from the assailant's throttling grip: discoid or confluent bruises from finger pads, abrasions from fingernails (crescent-shaped or scratches) and erythematous markings (**Fig. 3**). Nevertheless there are some victims who do not show any external evidence though there really has been an attempted

Table 2 Important external findings in physically injured victims of criminal assaults

Body region	*Findings*	*Kind of traumatization*
Head	Excoriation Bruising Laceration of the scalp or face	Blunt trauma to the soft tissues (e.g. kicks, blows from fists, flat hands or weapons, striking of the head against hard surfaces)
	Bleeding from the ear	Closed blows to the ear, fracture of the base of the skull
	Bleeding from the nose	Contusion or fracture of the nose, fracture of the base of the skull, severe strangulation
	Petechiae in the eyelids, conjunctivae and facial skin	Manual strangulation, ligature strangulation, traumatic asphyxia
	Bruising or laceration of the lips, loose or damaged teeth	Blow to the mouth
Neck	Discoid bruises, fingernail marks, scratches	Attempted manual strangulation
	Ligature mark	Attempted ligature strangulation
	Intradermal bruising from folds of cloth	Neck holds or locks
	Incised wounds	Sharp force (cuts, slashes, stabs)
Trunk	Abrasions, contusions	Blunt trauma
	Tram-line bruising	Impact of a rod or similar object
	Patterned abrasion or intradermal bruising	Impact of an object with a profile surface (e.g. rubber sole)
	Bite marks (two opposing bruises/abrasions) from the dental arches	Bites (child abuse, sexual assault)
	Brush abrasion ('grazes')	Tangential contact with a rough surface (caused by dragging)
	Penetrating wounds:	
	• Incised wounds	Sharp force (stabs)
	• Shot wounds	Projectile trauma
Limbs	Fingertip bruises (especially on the medial aspect of the upper arm)	Gripping or prodding
	Stab wounds and cuts (defense injuries)	Knife attacks
	Bruises and abrasions on the back of the hand and on the outer sides of the forearms (defense injuries)	Attack from blunt instruments, fists or feet
	Circumferential contusions of the wrists or ankles	Restraining by handcuffs, tying of hands or feet

strangulation: this may be true when the palm or forearm was placed over the neck and when a soft object, such as a garment or a pillow, was interposed during pressure. Prolonged gripping of the neck with concomitant occlusion of the cervical veins leads to congestion and petechial hemorrhages of the conjunctivae and on the facial skin, especially the eyelids (**Fig. 4**). Radiographs of the neck may reveal fractures of the hyoid bone and the thyroid cartilage; the incidence of such fractures is correlated with the extent of ossification mainly depending on the age and sex of the strangled person. Surviving victims mostly complain of pain on swallowing, on speaking and on neck movement. Other possible symptoms following attempted strangulation are unconsciousness and sphincter incontinence.

In ligature strangulation, the pressure on the neck is effected by a constricting object (telephone cord, nylons, stockings, belt, towel, scarf) tightened around the neck by a force other than the body weight. The appearance of the skin mark is influenced by the nature of the ligature (width, roughness, surface pattern), the force and duration of the strangulation, and the interaction with a resisting victim (displacement of the noose in attempts to remove it). The ligature mark usually encircles the neck horizontally (**Fig. 5**); in typical cases the skin reveals a transverse streak-like reddening caused by local erythema often associated with (partial) loss of the epidermis. Congestion and petechial hemorrhages above the ligature are usually more impressive than in victims of manual strangulation. On the other hand, damage to the hyoid or to the laryngeal cartilages is less common.

Suspected assailants may present examination findings indicative of possibly having been the offender in an attempted strangulation. The most frequent injuries seen in throttlers are nail marks inflicted by the opposing victim. Nail marks have been classified morphologically into three types: impression marks, claws and scratches. The majority of the skin lesions are located on the dorsal aspects of the forearms and hands; other classical sites are the face, neck, shoulders and anterior chest wall. Another typical kind of injury, especially seen in rapists, are bite

Figure 3 Neck of a rape victim with roundish bruises and scabbed abrasions from attempted manual strangulation.

marks on the hands. If a struggle takes place, the assailant may sustain nonspecific blunt trauma from blows or from wrestling with the victim.

Medical Hazards in Police Custody

Even in countries with a high standard of health care of detainees, some deaths do occur in police cells and in prisons. A large number of custodial deaths are due to natural causes, mainly from cardiovascular diseases, which account for the great majority of sudden and unexpected deaths. Another category of fatalities concerns suicides, mostly committed by hanging. In prisoners with suicidal tendencies it is therefore recommended to remove any objects that could be used for attempted strangulation like belts, ties, stockings and bootlaces. In spite of all precautions of the custodians, some detainees manage to find devices suitable for strangulation such as strips of bedding material or clothing.

Wrist cutting is another self-injurious behavior frequently seen in police custody, either as (attempted) suicide or as self-harm without intention to die. Incised wounds may be inflicted with any sharp-edged instrument such as a piece of glass or sheet metal from a tin. Risk factors associated with self-injurious behavior include the custodial setting itself (especially in isolation cells), being under the influence of alcohol and drugs at the time of incarceration, the availability of means for self-destructive behavior and many others.

A detainee who is suspected of being drunk and/or drugged at the time of committal needs particularly close observation. If the prisoner has consumed a large quantity of ethanol or (illegal) drugs just before arrest, he or she may be conscious and responsive at first, but he may become comatose later and die

Figure 4 Petechial hemorrhages on the skin of the eyelids and subconjunctival (scleral) hemorrhage in the medial angle of the left eye (one day after having been throttled).

Figure 5 Sharply defined ligature mark encircling the neck in a horizontal plane; the victim had been strangled with a shoe-lace three days before examination.

from acute alcohol or drug poisoning while he erroneously is thought to be sleeping off his drunkenness. This is not only true for ethanol ingestion but also for narcotic drugs (mostly heroin, methadone and codeine). The victims do not necessarily die from the depressive effects upon the brain (especially the respiratory center); in a high percentage, a secondary aspiration of vomit is found as the immediate cause of death.

Alcohol is an important causal factor in aggression and violent resistance frequently leading to the arrest of the suspect. Physical overpowering of an offender involves the risk of inflicting injuries. In other cases the person concerned may have sustained injuries due to falls or assaults before the police officers could intervene. It has to be stressed that even a fatal blunt trauma is not always associated with externally visible signs such as bruises, abrasions or lacerations. The authors had to give opinions on several custody deaths due to blunt head injuries without any wound or palpable swelling; the (deep) bruise on the site of impact became visible only when the inner aspect of the scalp was inspected during autopsy.

Severe and ultimately fatal head injuries need not be associated with unconsciousness in the early post-traumatic period. In persons developing an epidural or subdural hematoma, the symptoms due to the elevated intracranial pressure may set in only after a 'lucid interval' of several hours so that they are falsely thought to be uninjured. From this it follows that cases of suspected cranial trauma require close observation and the possibility of rapid transfer to a hospital.

Torture

According to the 1975 Tokyo Declaration of the World Medical Association, torture is defined as the deliberate, systematic or wanton infliction of physical or mental suffering by one or more persons acting alone or on the orders of any authority, to force another person to yield information, to make a confession, or for any other reason.

Torture has to be criticized as a particularly reprehensible offense against human rights, nevertheless it does exist in numerous countries throughout the world. Within the scope of this article, only a few medicolegal remarks on the physical manifestations of torture can be made. Unfortunately, often there is a long delay before medical examination so that conclusions as to the causation may be difficult because of the unspecificity of most scars.

Beating is probably the most common form of torture. Apart from kicks and blows from fists and flat hands, a great variety of weapons and instruments are used to inflict pain (for instance rifle butts, clubs and whips). When the body is struck by a rod or a similar object, each impact causes a double line of parallel bruises. Patterned abrasion or intradermal bruising may reflect characteristic details of the weapon's surface. Series of skin marks arranged in approximately the same orientation point to an unchanged position of the attacker in relation to the helpless (held or tied) victim. Heavy blows to the face are typically followed by peri-orbital hematomas ('black eyes'), extensive bruises and excoriations of the other facial regions, laceration of the lips and displacement of incisors, fractures of superficial bones (nasal bone, zygomatic bone, jaws). The other main targets of beating and whipping are the back, the buttocks, the legs and the soles of the feet, the abdomen, the breasts and the genitals which are also pinched and squeezed. Physical abuse other than blunt traumatization can only be mentioned briefly: cutting, piercing and stabbing, hair pulling, burning with cigarettes, repeated dipping of the victim's head under water, applying electric current, suspension and sexual abuse.

See also: **Causes of Death**: Blunt Injury; Sharp Injury; Burns and Scalds; Traffic Deaths. **Clinical Forensic Medicine**: Defense Wounds; Evaluation of Gunshot Wounds; Recognition of Pattern Injuries in Domestic Violence Victims; Self-inflicted Injury; Child Abuse; Sexual Assault and Semen Persistence. **Drugs of Abuse**: Antemortem Blood; Drugs and Driving. **Psychology and Psychiatry**: Psychiatry. **Accident Investigation**: Driver Versus Passenger in Motor Vehicle Collisions.

Further Reading

Bonneau J and Brown J (1995) Physical ability, fitness and police work. *Journal of Clinical Forensic Medicine* 2:157–164.

Buris L (1993) *Forensic Medicine*, pp. 293–311. Budapest: Springer.
DiMaio DJ and DiMaio VJM (1989) *Forensic Pathology*, pp. 87–102, 231–244. Amsterdam: Elsevier.
Forster B (ed.) (1986) *Praxis der Rechtsmedizin*, pp. 359–568. Stuttgart, New York: Thieme. (in German).
Härm T and Rajs J (1981) Types of injuries and interrelated conditions of victims and assailants in attempted and homicidal strangulation. *Forensic Science International* **18**:101–123.
Knight B (1991) *Simpson's Forensic Medicine*, 10th edn, pp. 65–86, 142–148, 222–225, 236–242. London: Arnold.
Knight B (1996) *Forensic Pathology*, 2nd edn, pp. 133–169, 295–303, 361–379. London: Arnold.
Ludwig S and Kornberg AE (eds) (1992) *Child Abuse. A Medical Reference*, 2nd edn, pp. 91–116, 175–190. New York: Churchill Livingstone.
Mason JK (1993) *Forensic Medicine. An Illustrated Reference*, pp. 52–71, 122–136, 153–165, 178–185. London: Chapman & Hall.
McLay WDS (eds) (1996) *Clinical Forensic Medicine*, 2nd edn. London: Greenwich Medical Media.
Pearn J (1989) Physical abuse of children. In: Mason JK (ed.) *Paediatric Forensic Medicine and Pathology*, pp. 204–220. London: Chapman & Hall.
Spitz WU (ed.) (1993) *Medicolegal Investigation of Death*, 3rd edn, pp. 199–235, 465–479. Springfield: C.C. Thomas.
Vanezis P (1989) *Pathology of Neck Injury*. London: Butterworth.

Child Abuse

R E I Roberts and **J V Evans**, The St Mary's Sexual Assault Referral Centre, St Mary's Hospital, Manchester, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0440

Introduction

Child abuse is a generic term which embraces a wide range of behavior engaged in by those in a position of trust with regard to a child. Such abuse may be perpetrated within or outside the family setting and the types of abuse the child is subject to overlap (**Table 1**). Child abuse has been present throughout history in all cultures, but although often ignored in the past, it has been increasingly recognized as a problem in recent years.

Table 1 Types of abuse

Physical (nonaccidental injury)
Sexual
Emotional
Neglect

The Role of the Physician or Child Care Professional

Everyone involved in the care of children has a responsibility to recognize the possibility of abuse and to recognize the primacy of the welfare of the child. Assessment and investigation of suspected abuse is a multidisciplinary task, involving welfare agencies, health professionals and the criminal and civil justice systems.

Recording the Findings

Health professionals have a duty to recognize the possibility of abuse and to document carefully all their findings. There is no substitute for detailed, accurate contemporaneous notes, which should include the initial complaint and any explanation given by the child or its carers verbatim.

An accurate description of any injury must be recorded, including details of its nature, size, position and age, using diagrams and photographs where possible. Where the child presents soon after the injury has been sustained consideration should be given as to whether the collection of samples for forensic examination would be useful, remembering that 'every contact leaves a trace' (Locard's principle).

In the preparation of reports for the courts and statutory authorities, it is of paramount importance that the health professional keeps an open mind and assesses any evidence objectively, considering all possible alternative explanations. A careful path must be trodden between over-interpreting the evidence on the one hand and failing to act decisively on the other where a child may be at risk. The emotional involvement of the professionals concerned has sometimes led to a lack of judgment and objectivity resulting in both over- and under-diagnosis of abuse.

Abuse usually escalates slowly over a long period of time, rather than being a single isolated event. However, a young or inadequate parent or carer may, in a fit of temper or frustration with an awkward, ill or disadvantaged child, lash out and cause serious injury, such as a skull fracture, shaking injury to the child or rupture to an internal abdominal organ, without really meaning to harm the child. At the other end of the spectrum sadistic abuse of the adult's power over a child is, unfortunately, not rare.

Often abuse is not recognized at the time of first injury, either because the child cannot or will not tell, or because carers and professionals do not consider the possibility of abuse. It is important to realize that there may have been previous injuries or even the death of a child or siblings whose abusive cause has not been recognized.

Physical Abuse

Nonaccidental injury (NAI) may range in its presentation from a few bruises and scratches to a dead child. The initial presentation (**Table 2**) may be to any agency involved with that particular child. It is important, therefore, that there is training and support in the recognition of possible abuse for those agencies. Factors which should alert the professional include a delay in seeking treatment, inadequate or discrepant explanations for the injury, the presence of injuries of different ages, a history of previous injuries, failure to thrive and a carer showing little or no anxiety about the child's condition.

Differential diagnosis

It is important to exclude other possible diagnoses, e.g. blood disorders or bone abnormalities, by taking a full medical history and completing a thorough examination. Every injury or pattern of injury should be evaluated so that the clinician can decide whether accidental or nonaccidental injury, carelessness, neglect or failure to cope, a traditional cultural practice or a medical condition offers the most satisfactory explanation for the clinical findings. The developmental stage of the child must be considered. A baby is not independently mobile and a clear explanation is required for any bruise or injury. However, it is important to recognize that bizarre accidents sometimes do happen and that they may not always be witnessed. Truth can sometimes be stranger than fiction.

The pattern of injury (**Table 3**) may be very helpful in determining causation and health care professionals should familiarize themselves with the typical patterns of fingertip and grip mark bruises, pinch marks and imprint abrasions and bruises left when a child is hit with an object, e.g. a belt, cord or stick. They should also be able to recognize innocent lesions which may mimic trauma, such as the Mongolian spot, juvenile striae or stretch marks and the marks left by coin rubbing (a traditional practice in some cultures). The aging of injuries is also very important but may be fraught with difficulty, especially with regard to the dating of bruising.

Deliberately inflicted injury may be such that fracture of the bones occurs. It is for this reason that a full skeletal survey is frequently indicated (**Table 4**) when physical abuse is suspected, especially in the very young child. The interpretation of skeletal injuries in suspected child abuse must be approached with great caution and the advice of a senior radiologist with expertise in the field of child abuse sought.

The head is the commonest target for assault in the young child and head injury is the major cause of death after physical abuse. Most (95%) serious intracranial injury in the first year of life is the consequence of NAI. Such injury occurs in one of two ways: either, as a result of direct impact trauma, such as punching to the head or throwing or swinging the child so that the head comes into contact with a hard object; or, as a result of diffuse brain injury due to the acceleration–deceleration of shaking. This latter type of injury is usually associated with retinal hemorrhages.

Skull fractures sustained in accidental circumstances are typically single, narrow, hairline fractures, involving one bone only, usually the parietal, and are not

Table 2 Presentation

- Delay in seeking treatment
- Inadequate or changing explanation
- Lack of any explanation
- Injuries of different ages
- History of previous injury
- Failure to thrive
- Parent shows little concern
- Frozen awareness in child

Table 3 Common patterns of injury

- Any bruising on a baby
- Multiple bruises
- Pattern bruises
- Finger marks/hand weals
- Bilateral black eyes
- Torn upper frenulum
- Bite marks
- Cigarette burns
- Scalds and burns
- Fractures and head injury

Table 4 Indications for skeletal survey

1. The history, type or pattern of injury suggests physical abuse.
2. Older children with severe bruising.
3. All children under 2 years of age where nonaccidental injury is suspected.
4. When a history of skeletal injury is present.
5. In children dying in suspicious or unusual circumstances.

often associated with intracranial injury. Depressed fractures can occur but are localized with a clear history of a fall onto a projecting object. In contrast, skull fractures which arise following NAI are typically multiple, complex and branched. The fractures are wide and often grow. Occipital fractures are highly specific for abuse but bilateral fractures occur, and often more than one bone is involved. A depressed fracture may occur singly as part of a complex fracture or there may be multiple, extensive such areas. Severe associated intracranial injury is common.

Rib fractures are usually occult and detected only on X-ray or by radionuclide bone scanning. They are seen most commonly in infants and young children. In infants, in the absence of bone disease, the extreme pliability of the ribs implies that such fractures only occur after the application of considerable force.

Epiphyseal and metaphyseal fractures are injuries typical of NAI. The pulling and twisting forces applied to the limbs and body of a small child when it is shaken lead to disruption of the anatomy of the shafts of the long bones, subsequent periosteal reaction and ultimately new bone formation. Near the joints small chips of bone or the entire growth plate (corner and bucket handle fractures) may become separated from the shaft of the bone. Several different types of fracture of different ages and in different sites may be present in the same child.

Burns and scalds are also commonly seen in physical abuse. Again it is important to ascertain whether the explanation for the injury is consistent with the pattern of injury actually observed. The sharply demarcated area of tissue damage to the buttocks in the child who has been held in very hot water, for example, is quite different from the relatively haphazard pattern of injury in a child who has accidentally fallen into a container of hot water or tipped it over themselves.

Cigarette burns typically form a circular or oval cratered lesion 0.6–1.0 cm across which heals with scarring because it involves the full thickness of the skin. Causing such an injury is a sadistic act which involves the burning end of the cigarette being held in contact with the skin for 2–3 s. Accidentally dropping a cigarette on to a child or the child brushing against a lighted cigarette will cause only a very slight superficial burn.

Munchausen's syndrome by proxy is a condition in which a carer, usually the mother, deliberately invents stories of illness in her child and then seeks to support this by causing appropriate physical symptoms. This often leads to the child being subjected to many unnecessary often invasive investigations and may involve the mother deliberately harming her child, for example by administering poisons or salt or by smothering the child. Sometimes such actions may lead to the death of the child.

Sudden infant death syndrome is the term applied to infants who die without their cause of death being ascertained. Many explanations for this phenomenon have been put forward, including, controversially that a small number of these deaths may in fact be undetected deliberately inflicted asphyxia.

Child Sexual Abuse

Child sexual abuse is any use of children for the sexual gratification of adults.

This may occur both inside and outside the family setting but always involves a betrayal of the child's trust by the adult and an abuse of power. It encompasses a very wide range of behavior on the part of the adult, much of which will have a psychological effect on the child but may not leave any discernible physical evidence. Medical evidence, if present, is often equivocal, and forensic evidence, such as semen, is a rare finding, particularly in view of the fact that such abuse does not usually present to professionals acutely.

The child who may have been sexually abused may present in a number of ways. A spontaneous allegation by a child must always be taken very seriously. However, recent work has confirmed the child witness's vulnerability to suggestion, a vulnerability which may lead to honest mistakes on the part of the child and also leave them open to exploitation by adults with their own agendas. Some children may present with sexualized behavior or with nonspecific signs of emotional disturbance or disturbed behavior. The child may be living in an environment where the risk of sexual abuse is thought to be high or there may be medical findings suggestive of sexual abuse.

The medical examination for child sexual abuse

It is important that the clinician is clear as to why examination of the child's genitalia is necessary (**Table 5**). In every case it is good practice to seek formal written consent to examination from the adult who has parental responsibility for the child as well as explaining to the child themselves what is happening and why. Where a child withdraws consent for any part of the examination this withdrawal should be acknowledged and respected by the examining clinician. Loss of control over what is done to them is an

Table 5 Reasons for medical examination

Forensic	*Therapeutic*
• Establish cause of injuries	• Diagnose and treat injuries
• Take valid specimens: semen saliva fibers	• Check for medical causes • Sexually transmitted diseases • Pregnancy • Reassurance Start recovery

important feature in abuse and this should not be compounded by the medical or forensic process.

In suspected child sexual abuse the genitalia should be examined: to detect injuries, infection or disease which may need treatment; to reassure the child (and the carers) who may often feel that serious damage has been done; and, hopefully to start the process of recovery. In addition to these therapeutic reasons there are medicolegal reasons why examination is important: to help evaluate the nature of the abuse, and possibly to provide forensic evidence. It is important that the child is not abused by the process of examination itself and this should be well planned and kept to a minimum. A single examination whether carried out by one clinician alone or as a joint examination with a colleague is ideal and must be sensitive to the needs of the child.

The external genitalia should be carefully inspected to assess genital development and for any evidence of abnormality or injury. It is important that skin disease, such as lichen sclerosus et atrophicus, or infection, is recognized and that any changes produced are recognized as such and not mistakenly attributed to trauma.

The hymen should be carefully examined, looking for notches, bumps and healed tears. Typically blunt force penetrating trauma, such as penile penetration, causes a full width tear of the hymen which may extend posteriorly into the fourchette. The quality of the edge of the hymen is important, as is the presence or absence of attenuation (loss of tissue). The size of the hymenal opening should be assessed although unless this is grossly enlarged (more than 1.5 cm diameter in the prepubertal child), its significance is debatable.

Where an abnormality of the hymen is suspected it is important that this is shown to be real and not due to examining position or poor technique. This can be established by demonstrating persistence of the feature by examining the child in more than one position, e.g. supine frog leg and knee chest, or using a technique to display the edge of the hymen, e.g. using a moistened swab or glass rod.

Interpreting the medical findings

At the end of the assessment the examining clinician should ask himself/herself whether the findings are normal or abnormal, and, if the latter, why? What has caused the abnormality? Is it abuse or not? Could the explanation be one of congenital anatomical variation, disease, accidental trauma or nonpenetrative or penetrative deliberately inflicted trauma. The clinician should carefully consider what weight should be attached to any findings, for example are they nonspecific, supportive but not diagnostic, or diagnostic of blunt force penetrating trauma, such as digital or penile penetration, through the hymen.

Positive medical findings are only present in a minority of cases. This is because much sexual abuse is of a type which does not lead to physical damage and victims often do not come to attention before any minor injuries have healed. Also there is a wide range of normal anatomical variation in the structures in the genital and anal region and studies have shown that many of the findings thought to be suggestive of sexual abuse are in fact present in the nonabused population. It is important, therefore, that the examining clinician is familiar with normal anatomy (**Fig. 1**) and the changes which take place in its appearance with age (from infancy through the

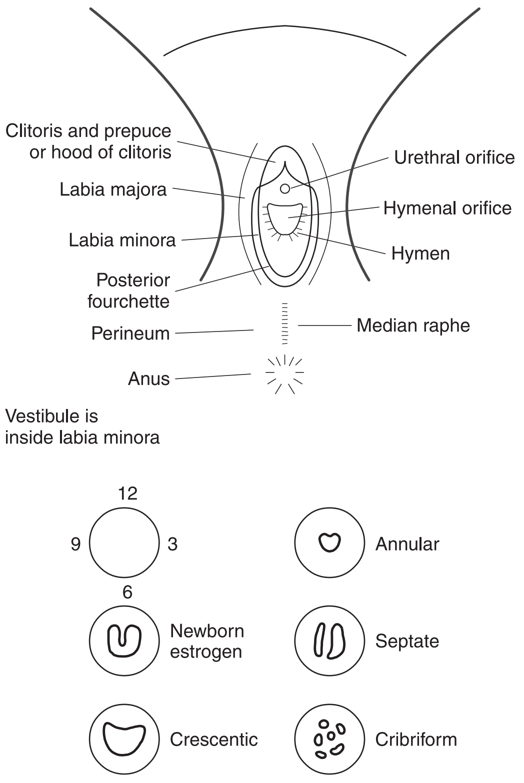


Figure 1 The prepubertal vulva.

prepubertal years to puberty), is able to describe any findings accurately using a standard nomenclature, and, where possible, record the findings photographically so as to allow peer review of the findings.

Intercrural intercourse (pushing the penis between the thighs against the genitalia with little or no actual penetration) usually does not result in any injury although occasionally there may be bruising to the perineum. Nonspecific findings such as patchy redness of the labia and perineum are common. Splits to the fourchette do occur and may sometimes heal with scarring. Pressure by the penis on the outer genitalia is perceived as painful by the child who may be honestly mistaken as to whether penetration has taken place or not.

Accidental trauma, such as straddle injury, usually causes injury to the vulva, fourchette or vestibule. Such injuries are commonly anterior, unilateral and consist of bruising of the labia, often with a deep bruise or even a split in the cleft between the labia majora and minora or between the labia minora and the clitoris. The hymen is an internal structure protected by the vulva and will only be damaged if there is a penetrative component to the injury.

Anal findings

The significance of anal findings or their absence has been a much debated subject. Most, such as bruising, abrasion, fissures, 'reflex' anal dilatation and anal laxity, can be caused by penetrative abuse or may have an innocent explanation. Where lubrication has been used there may be no positive clinical findings even when a child is examined within hours of penile anal penetration having occurred. The only diagnostic finding is semen in the anal canal or rectum and this is rare.

Fissures occur when the anal tissues are dilated beyond their elastic limit. They should be carefully distinguished from prominent folds in the anal canal. They may be multiple or single. They are acutely painful at the time of their formation and there is usually some associated bleeding which can vary from the trivial to the torrential. However, other causes of bleeding should be sought. Fissures, especially if they become chronic, may heal with scarring and the formation of skin tags.

The skin of the anal verge may be swollen after recent abuse or may become thickened, rounded and smooth after repeated abuse but such changes are not sufficient in themselves to lead to suspicions of abuse.

Provided that the bowel is normal and that there is no neurological disorder, abnormal anal sphincter tone implies that something hard and bulky has passed through the sphincter, but does not distinguish the direction of movement, from inside out or vice versa. Any history of medical investigation, treatment or bowel problems is, therefore, likely to be of significance and should be actively sought, both from the child and the carer.

The internal anal sphincter maintains anal closure and continence. Stretching or tearing of the muscle fibers may lead to incompetence of the sphincter and to anal soiling. The presence of feces in the lower rectum may also lead to physiological relaxation of the sphincter.

'Reflex' anal dilatation is not a reflex at all, but refers to the opening of the anal sphincter some 10 s after gentle buttock separation has been applied. It may be present in both abused and nonabused children and in the presence of feces in the rectum. Its significance is unproven and although it may raise the suspicion of abuse it is not a reliable sole diagnostic sign.

Forensic evidence

When children present acutely after suspected sexual abuse, consideration should always be given to the collection of forensic samples both for biological (saliva, semen, hairs etc.) and nonbiological trace evidence (fibers).

Where children have been subjected to sexual abuse by strangers then tests for sexually transmitted diseases should always be considered, discussed with the carers and taken where possible. In intrafamilial cases it is desirable to take tests from the child but, especially where this may cause distress, asking other members of the family, such as the suspected perpetrator, to attend the genitourinary clinic may be helpful. Where infection is present it is desirable to seek the opinion of a specialist in genitourinary medicine with regard to treatment and the medicolegal significance of that particular infection. This involves assessing such factors as the age of the child, the prevalence of the disease in the community, its presence or absence in the alleged perpetrator, its incubation period and possible modes of transmission.

Genital warts in a child can be transmitted by sexual contact and should raise a suspicion of sexual abuse. However, warts have a long incubation period and transmission during the birth process and from the mother or carer during toileting procedures should be considered. Self-inoculation or transmission from innocent skin contact may lead to genital or anal wart infection with the common wart virus.

In postpubertal girls the possible need for postcoital contraception or a pregnancy test must be addressed.

Emotional Abuse and Neglect

Emotional abuse and neglect of children may take many forms, including failure to meet physical needs, a failure to provide consistent love and nurture through to overt hostility and rejection. It is rare for other forms of abuse to have occurred without some element of emotional abuse at least.

The deleterious effects on children are manifold and are very much age dependent as to how they present. In the infant the presentation is usually with failure to thrive, recurrent and persistent minor infections, frequent attendances at the Accident and Emergency Department or admissions to hospital, severe nappy rash or unexplained bruising. There is often general developmental delay and a lack of social responsiveness. These signs and symptoms may have alternative innocent organic explanations and these must always be excluded.

A preschool child who has suffered emotional abuse and neglect will often present as dirty and unkempt and of short stature. Socioemotional immaturity, language delay which may be profound, and a limited attention span may also be displayed. This may be associated with aggressive, impulsive and often overactive behavior which may lead to increasing hostility and rejection from their carers. However, behavior with strangers is in direct contrast to this as the child may be indiscriminately friendly, constantly seeking physical contact ('touch hunger') and reassurance.

In the school age child, persistent denigration and rejection leads to very low self-esteem, coupled with guilt as the child feels responsible for the way things are and self-blame is common. This leads to a child with few or dysfunctional relationships who may exhibit self-stimulating behavior or self-harm. Disturbances in the pattern of urination or defecation are common. Sometimes behavior may be bizarre and extreme. Physically the child may be of short stature, be unkempt and have poor personal hygiene. Learning difficulties, poor coping skills and immaturity of social and emotional skills may also be exhibited.

In the adolescent, although emotional abuse and neglect is less often reported, it is common and may start at this age especially in girls. It can have a profound effect on adolescent adjustment and usually results in a whole range of self-harm and behavioral problems both contemporaneously and into early adult life. These include sexual promiscuity, early pregnancy, running away, suicide attempts, self-harm, drug and alcohol abuse, and delinquent behavior. Intervention in this age group is fraught with difficulty and is often the most neglected, reinforcing the adolescent's perception of rejection and low self-esteem.

Where failure to thrive and short stature are an issue, it is important that organic causes are excluded and corrected before this is presumed to be due to emotional abuse and neglect. There is no substitute for regular and accurate height and weight measurements which should be taken and plotted on the appropriate centile chart whenever a child presents with a possible diagnosis of child abuse. Where abuse is a factor, such charts will show a characteristic catch up growth when the child is removed from the source of abuse. Disordered patterns of eating and behavior relating to food will often resolve at this time also.

In conclusion, adults can behave in a wide variety of ways which result in the abuse of their power over children. Such abuse may have emotional, behavioral and physical consequences which in extreme cases may lead to the death of a child. Working with survivors of all types of abuse to minimize its effect on their lives and enable such children to fulfill their potential is a task which must be shared by professionals and society at large. Only then can the intergenerational cycle of abuse and deprivation be broken. The role of the forensic clinician is to bring together professional skills and knowledge about the nature of injuries, the possible mechanisms of their causation, patterns of healing and variations in appearance with time, and an awareness of appropriate forensic investigations, such as the careful gathering of samples of body fluids, trace evidence and radiographs. With the welfare of the child as the prime concern, and treating all the parties involved with consideration and respect at all times, the clinician must use these skills to work as part of a multidisciplinary team in the interests of the child.

See also: **Causes of Death**: Burns and Scalds. **Clinical Forensic Medicine**: Sexual Assault and Semen Persistence, **Deoxyribonucleic Acid**: Significance. **History**: Forensic Sciences. **Legal Aspects of Forensic Science**. **Anthropology**: Skeletal Trauma.

Further Reading

Bays J and Chadwick D (1993) Medical diagnosis of the sexually abused child. *Child Abuse and Neglect* **17**:91–110.

Ceci SI and Bruck M (1993) Suggestibility of the child witness. A historical review and synthesis. *Psychological Bulletin* **113**:403–439.

Emans SJH and Goldstein DP (1990) *Paediatric and Adolescent Gynaecology*, 3rd edn. Boston: Little, Brown.

Heger A and Emans SJ (1992) *Evaluation of the Sexually Abused Child*. New York: Oxford University Press.

Langlois NEI and Gresham GA (1991) The ageing of bruises: a review and study of the colour changes with time. *Forensic Science International* **50**:227–238.

Lynch PJ and Edwards L (1994) *Genital Dermatology*. Singapore: Churchill Livingstone.

McLay WDS (ed.) (1996) *Clinical Forensic Medicine*, 2nd edn. London: Greenwich Medical Media.

Meadow R (ed.) (1997) *ABC of Child Abuse*, 3rd edn. London: British Medical Journal.

Robinson SP (1996) *Principles of Forensic Medicine*. London: Greenwich Medical Media.

Root I (1992) Head injuries from short distance falls. *American Journal of Forensic Medicine and Pathology* **13**:85–87.

Royal College of Physicians (1997) *Physical Signs of Sexual Abuse in Children*, 2nd edn. Report of a Working Party. London: Royal College of Physicians.

Defense Wounds

S Pollak, University of Freiburg/Br, Freiburg, Germany
P J Saukko, University of Turku, Turku, Finland

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0438

Introduction

The presence of defense wounds is justly considered an important sign that an individual has been the victim of an assault. Another aspect of medicolegal significance concerns the physical activity of the victim: defense wounds indicate that the attacked person was – at least initially – conscious and able to use his limbs for protection resulting in injuries to the forearms or hands or – in rare cases – to the legs. When sustaining defense wounds, the victim must have been aware of the assault (anticipating the blow or thrust) and fit for the natural reaction of protecting himself.

Defense Wounds from Knife Attacks

Such injuries are seen when the victim attempts to ward off a knife either by seizing it or by raising forearms and hands in order to protect other body regions (head, neck, chest).

Traditionally a distinction is made in the literature between active and passive defense wounds. According to this distinction *active* defense wounds occur when the victim grasps the knife with the hand with the injury then being located on the palmar side of the hand. The *passive* wound type is sustained when the victim raises their hands or arms to protect the attacked body region; in this case the injuries will primarily be located on the extensor side.

There is only a small number of papers available with detailed statements as to the frequency and distribution of defense wounds. The frequency of such wounds indicated for victims killed by sharp force usually ranges between 30% and 50%. In victims primarily fit for reaction the probability of the presence of defense wounds increases with the number of hits placed on the body.

If in spite of multiple stabs, no defense injuries are found, it has to be considered that the victim may have been unable to react or impaired in his ability to defend himself already before the use of sharp force: for example in knife attacks on a sleeping person, in unexpected assaults and in victims made unconscious before the attack. The absence of defense wounds is also observed when the victim was held or tied. The same also largely applies in those cases where severe traumatization of a different type (e.g. by blunt force) occurred before the use of sharp force. Moreover, it is to be expected that in cases of very high blood alcohol concentrations or under the influence of drugs the ability to defend oneself adequately against an assault is reduced.

A study of large numbers of cases with regard to the localization of defense wounds found about the same number of injuries on the flexor and the extensor side, often combined in the same victim. It seems justified to express doubts as to the stereotyped distinction between active and passive defense wounds. First, the various possible ways stabbing with a knife must be taken into consideration. Different ways of moving the knife (in a downward or an upward direction) may produce defense wounds on either the extensor or flexor side for the same position of the arm. An injury on the extensor side of the limbs may also occur if the victim tries to seize the knife and the blade comes into contact with the back of the hand or the extensor side of the forearm.

In an attempt to grasp the knife the fingers of the victim may close around the blade, which cuts across the flexures of the phalanges as the knife is withdrawn (**Fig. 1**). In this case several fingers may be injured by a single impact of the cutting edge sometimes even completely severing the flexor tendons. Another typical location is the web between the thumb and the index finger. Incised wounds of the hands are often 'shelved' so that one margin forms a flap of skin (**Fig. 2**). Hand injuries are frequently located at the transition from the flexor to the extensor side of the fingers. Such wounds are typically encountered if the victim tries to protect himself by holding his hand against the knife with spread fingers.

Figure 1 Two cuts on the palmar surface of the left hand caused by the cutting edge of a machete when the victim attempted to clutch the weapon.

It should be mentioned that more than two-thirds of all defense wounds are found on the left arm or left hand. To understand this phenomenon one has to visualize the position of the attacker and the victim: if somebody is in a threatening situation he will first try to bring his body into an upright position. Thus a lying person will try to stand up, a sitting person will get up. As soon as the attacker and the victim are standing opposite each other, the victim will try to ward off the assaulting hand. In doing this he will mostly rise his left arm, because it is nearest to the attacking arm of a right-handed assailant. The arm raised at the shoulder joint and flexed at the elbow is mostly hit by the assailant's weapon on the extensor side.

Figure 2 Defense wound of the right ring finger with beveled margin forming a skin flap.

In surprise attacks the described motion of the victim, which refers to the primary phase of the attack, is largely unconscious and very fast. Another reason why more injuries are located on the left-hand side may also be the unconscious attempt of the victim to maintain the function of the right arm (which is the 'stronger' arm in most persons) and thus to ward off the assailant with the often weaker left arm. In this connection it is of course important, whether the assailant and the victim are right- or left-handed. In rare cases defense wounds are localized even on the legs; if the victim was in a lying position he may have used his legs to ward off the knife or curled up in order to protect vital areas with his drawn up extremities.

Stabbing actions do not necessarily result in actual stab wounds but also in cuts if the blade contacts the curved surface of a limb tangentially (**Fig. 3**). In transfixion stab wounds the weapon passes through the soft tissues of an extremity and afterwards possibly re-enters another body region (**Fig. 4**).

In most cases there is considerable relative movement between the assailant and the victim. According to the dynamics of such knife attacks, an irregular wound configuration is often found. When the knife has been plunged into the body, twisting of the blade before withdrawal or turning of the victim causes wounds shaped like the letters L, Y or V (**Fig. 5**).

In cuts and stab wounds, shelving is likely if the blade enters at an oblique angle so that one margin is beveled and the other one undermined. Stab wounds associated with an imprint abrasion ('hilt mark') indicate that the blade was vigorously pushed in up to its end; a patterned abrasion possibly corresponds to the respective construction parts of the knife (e.g. handle or guard). In victims of knife attacks, incised stab wounds due to the combined action of cutting

Figure 3 Multiple cuts of the left forearm and hand sustained when trying to ward off a knife attack against the neck.

Figure 4 Transfixion stab wound of the left hand after surgical treatment. The victim tried to fend off an attack with a kitchen knife.

Figure 6 Defense injuries on the forearm of a victim who was killed by stab wounds to the chest and cut-throat wounds.

and stabbing elements are a frequent finding (**Fig. 6**). The wound may start as a cut which terminates as a stab wound; on the other hand some stab wounds turn into an incised wound as the knife is withdrawn at a shallow angle.

Defense injuries to limbs covered by garments are usually accompanied by corresponding damages/cuts in the clothing, whereas a large majority of the suicides bare the target regions before cutting or stabbing.

Defense Wounds Due to Blunt Force

Defense wounds of this kind are seen in blunt force attacks when the victim attempts to ward off kicks, blows of the fist, or blunt instruments. Their location is similar to defense wounds from knife attacks (hands, forearms or, less frequently, the legs). Injuries inflicted to the thighs possibly result from attempts to shield the genitals against blows or kicks aimed at the lower part of the body.

The injuries most often seen are abrasions and contusions (with concomitant bruising) on the back of the hands, wrists and forearms (**Fig. 7**). Though abrasions and bruises are the most common defense injuries due to blunt force, additional lacerations may occur in skin regions supported by bone (for example the metacarpal region and the knuckles). Lacerations from a blunt instrument possibly contain embedded foreign material indicative of the weapon used. Defense injuries inflicted together with wounds in the regions the victim wanted to protect with his hands and arms show the same stage of healing on later inspection.

In rare cases the victim may sustain fractures of the hand bones. Another classical fracture site is the ulnar shaft, which is exposed in victims parrying a blow or

Figure 5 Irregular stab wounds on the dorsal and ulnar aspect of the right forearm: one of the stab wounds is extended by a superficial incision, the other one is V-shaped due to twisting of the blade. The weapon used by the assailant was a survival knife.

Figure 7 Bruises on the dorsal face of the left hand of a female victim who tried to protect her head against fist blows. In addition there is a shallow cut on the index finger (defense wound from a knife attack).

kick with their raised forearm. Depending on the site of the causal impact these fractures usually occur at the transition from the proximal to the middle third or in the middle third of the ulna.

Defense wounds due to blunt trauma deserve special attention in suspected physical child abuse. In such cases bruises or abrasions may reflect patterned details from the instrument delivering the blow (e.g. belt, stick, rod, heel of a shoe). In adults also a great variety of blunt instruments can be used to inflict severe injuries, mainly to the head. If the hand is interposed in a protecting manner, it will primarily be hit by the weapon (**Fig. 8**). By placing the hands on the top of the head the victim tries to lessen the blows. Apart from typical blunt weapons as hammers, firm bottles and wrenches, sometimes also (blank cartridge) pistols and revolvers are used for striking ('pistol whipping').

Figure 8 Defense injuries on the back of the right hand: excoriations, bruises and a shallow laceration due to multiple blows with a pistol grip.

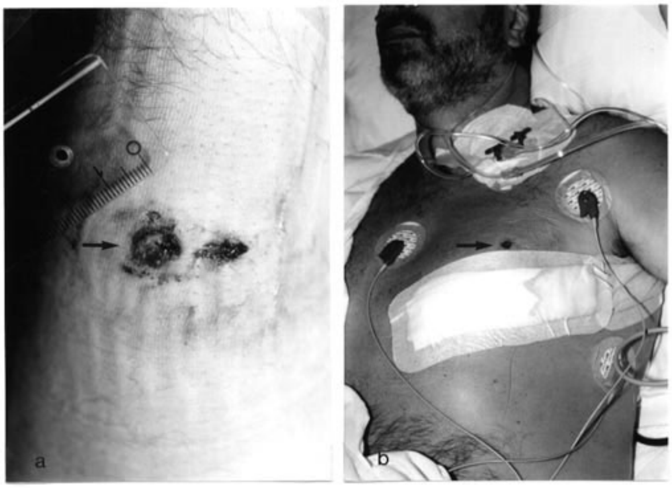
Defense Wounds from Other Weapons

Chop wounds are a special category of injuries due to sharp force by heavy instruments with a cutting edge like axes, machetes and meat cleavers. Such wounds are characterized by cuts of the skin in combination with sharp-edged notches or comminuted fractures of the underlying bone. In attempts to ward off an assault with a chopping weapon, the back of the hand sustains deep chop wounds. Finally some uncommon stabbing instruments have to be mentioned: icepicks, forks, files, pens, scissors, screwdrivers, broken bottles.

In a wider sense even in firearm injuries the upper limbs may reveal a special kind of defense wound when the victim raises an arm in an unsuccessful attempt to shield vital areas of the body. Under such circumstances the bullet first passes through the hand or the arm and afterwards often enters the head or the thorax producing a re-entry wound (**Fig. 9**).

Misleading Findings Similar to Defense Wounds

Superficial cuts restricted to the skin are sometimes found on the fingers of suicides who used double-edged razor blades for cutting their wrists. Most of these cuts do not pass through all layers of the skin so that differentiation from real defense wounds should not be difficult.



A B

Figure 9 Entrance and exit wound of a caliber 0.32 bullet which first passed through the right forearm (**A**) and afterwards entered the chest (**B**). The victim had raised his arm when a burglar shot at him.

If a perpetrator stabs his victim using a knife without a guard, his bloody hand may slip from the handle onto the blade producing cuts on the palm or the fingers, which at first sight give the impression of real defense wounds. As a consequence, blood stains of both the victim and the assailant might be detected on the blade.

Heavy blows with the fist sometimes result in injuries to the assailant's hand, for instance bruises or wounds from the victim's incisors or even fractures of the metacarpal bones.

The features of self-inflicted injuries are dealt with in a separate article. Hands and arms may bear multiple incisions from a preliminary suicide attempt, mostly localized across the front of the wrists and on the flexor surface of the forearms. The typical appearance is characterized by multiple clustered wounds, mainly in a horizontal and parallel arrangement, some being very superficial tentative incisions ('hesitation cuts').

Apart from persons with a real suicidal tendency, numerous incisions are also seen among patients suffering from psychiatric disorders and in persons who want to draw attention to themselves ('sympathy cuts') particularly by making a false accusation of assault.

See also: **Causes of Death**: Sharp Injury. **Clinical Forensic Medicine**: Self-inflicted Injury.

Further Reading

Buris L (1993) *Forensic Medicine*, pp. 51, 59, 63, 71. Budapest: Springer.

Crane J (1996) Injury. In: McLay WDS (ed.) *Clinical Forensic Medicine*, 2nd edn, pp. 143–162. London: Greenwich Medical Media.

DiMaio DJ and DiMaio VJM (1989) *Forensic Pathology*, pp. 102, 195–199. New York: Elsevier.

Knight B (1991) *Simpson's Forensic Medicine*, 10th edn, pp. 84–85. London: Arnold.

Knight B (1996) *Forensic Pathology*, 2nd edn, pp. 162–163. London: Arnold.

Kornberg AE (1992) Skin and soft tissue injuries. In: Ludwig S and Kornberg AE (eds) *Child Abuse*, 2nd edn, pp. 91–104. New York: Churchill Livingstone.

Mason JK (1993) *Forensic Medicine. An Illustrated Reference*, pp. 64, 128. London: Chapman & Hall.

Metter D and Benz D (1989) Defence-injuries caused by sharp force in homicides. *Zeitschrift für Rechtsmedizin*, **102**:277–291 (in German).

Prokop O and Radam G (1992) *Atlas der gerichtlichen Medizin*, 3rd edn. pp. 104, 240, 247, 249, 252, 275, 277, 328, 356. Berlin: Ullstein Mosby (in German).

Spitz WU (1993) Sharp force injury. In: Spitz WU (ed.) *Medicolegal Investigation of Death*, 3rd edn, pp. 252–310. Springfield: CC Thomas.

Spitz WU and Platt MS (1993) The battered child and adolescent. In: Spitz WU (ed.) *Medicolegal Investigation of Death*, 3rd edn, pp. 687–732. Springfield: CC Thomas.

Watson AA (1993) Stabbing and other incisional wounds. In: Mason JK (ed.) *The Pathology of Trauma*, 2nd edn, pp. 97–108. London: Arnold.

Evaluation of Gunshot Wounds

W S Smock, University of Louisville School of Medicine, Louisville, KY, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0750

Introduction

Emergency physicians in the United States treat in excess of 250 000 gunshot wounds per year. More than 90% of these wounds are the result of projectiles discharged from handguns. Historically, emergency physicians have been trained to treat gunshot wound injuries without consideration for the forensic issues associated with them. As a consequence, emergency department personnel have misinterpreted wounds, have inadvertently altered their physical characteristics, and destroyed critical forensic evidence associated with the wounds in the process of rendering patient care. These problems are avoidable. Emergency room personnel should have the tools and knowledge base necessary to determine the following forensic issues: range of fire of the offending weapon (projectile) and entrance wound versus exit.

Misdiagnosis and Errors of Interpretation

The majority of gunshot wound misinterpretations result from the fallacious assumption that the exit wound is *always* larger than the entrance wound. The size of any gunshot wound, entrance or exit, is primarily determined by five variables: the size, shape, configuration and velocity of the bullet at the instant of its impact with the tissue and the physical characteristics of the impacted tissue itself. When the exiting projectile has substantial velocity or has fragmented, or changed its configuration, the exit wound it generates may in fact be larger than the entrance wound associated with it. Conversely, if the velocity of the exiting projectile is low, the wound it leaves behind may be equal to or smaller than its corresponding entrance wound. If the kinetic energy of the bullet is

transferred to underlying skeletal tissue, bone fragments may be extruded from the exit wound, contributing to the size and shape of the wound. The elasticity of the overlying connective tissue and skin can also affect wound size. Misinterpretations of gunshot wounds by nonforensically trained physicians have ranged as high as 79%. To avoid the misinterpretation of wounds, health practitioners who lack forensic training should limit their documentation of the wounds to detailed descriptions of their appearance and not attempt to make the determination: 'entrance or exit'.

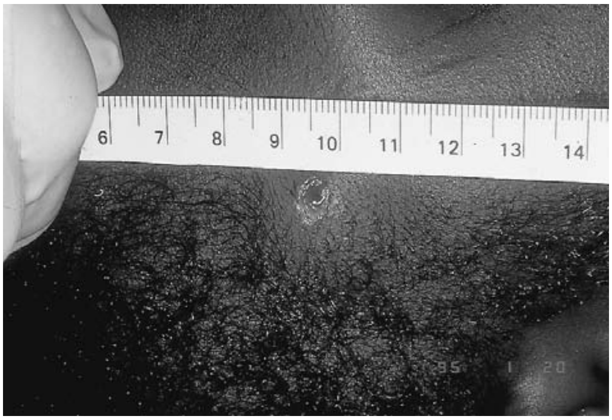


Figure 2 The area of abraided tissue surrounding this gunshot wound of entrance is the 'abrasion collar'. Additional appropriate terms would include: abrasion margin, abrasion rim or abrasion ring.

Entrance Wounds

Entrance wounds can be divided into four general categories, dictated by the range of fire from which the bullet was discharged. Range of fire is an expression used to describe the distance between the gun's muzzle and the victim/patient. The different ranges of fire are: distant or indeterminate range; intermediate or medium range; close range and contact. The entrance wounds associated with each of the four ranges of fire categories will have physical characteristics pathognomonic to each category.

Distant or indeterminate entrance wounds

The distant or indeterminate gunshot wound of entrance is one which results when the weapon is discharged at such a distance that only the bullet makes contact with the victim's skin. When the bullet or projectile penetrates the epithelial tissue, there is friction between the skin and the projectile. This friction results in an abraded area of tissue which surrounds the entry wound and is know as an abrasion collar (**Figs 1** and **2**). In general, with the exception of gunshot wounds on the soles of the feet and palms of the hands, all handgun gunshot wounds of entrance will have an associated abrasion collar. The width of the abrasion collar will vary depending on the caliber of the weapon, the angle of bullet impact and the anatomical site of entrance. Skin which overlies bone will generally have a narrower abrasion collar than skin supported by soft tissue. Entrance wounds on the soles and palms are usually slit-like in appearance. It is important to note that the abrasion collar is not the result of thermal changes associated with a 'hot projectile'. The terms abrasion margin, abrasion rim, abrasion ring and abrasion collar are all used interchangeably.

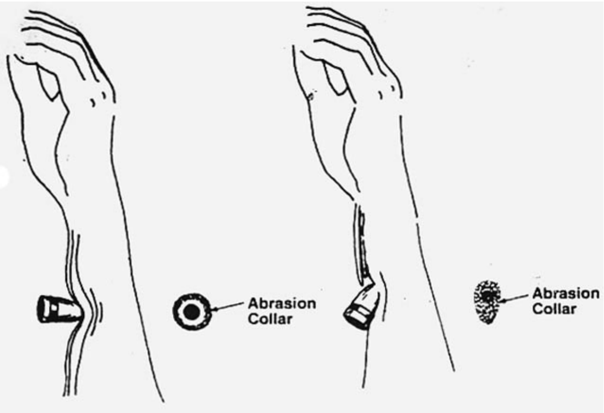
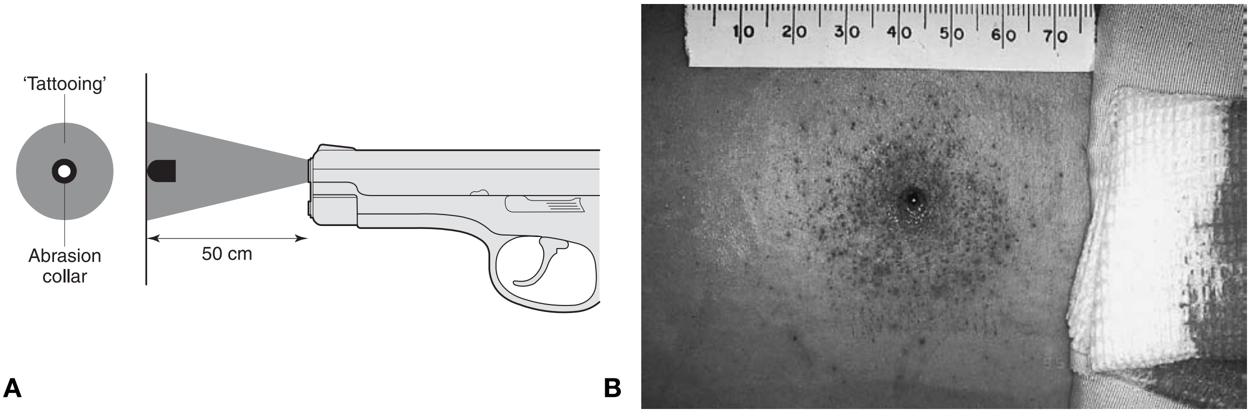


Figure 1 An abrasion collar is the abraided area of tissue surrounding the entrance wound created by the bullet when it dents and passes through the epithelium. The abrasion collar will vary with the angle of impact.

Intermediate range entrance wounds

'Tattooing' is pathognomonic for an intermediate range gunshot wound. Tattooing is a term used to describe the punctate abrasions observed when epithelial tissue comes into contact with partially burned or unburned grains of gunpowder. These punctate abrasions cannot be wiped away and will remain visible on the skin for several days. Clothing, hair or other intermediate barriers may prevent the powder grains from making contact with the skin. Though it has been reported, it is rare for tattooing to occur on the palms of the hands or soles of the feet, because of the thickness of the epithelium in these areas.

The density of the tattooing will be dictated by: the length of the gun's barrel, the muzzle-to-skin distance, the type and amount of gunpowder used and the presence of intermediate objects. Punctate abrasions from unburned gunpowder have been reported with distances as close as 1 cm, and as far away as 100 cm (**Fig 3A**, **B**).



Figures 3A, B 'Tattooing' is the term used to describe punctate abrasions from the impacting of unburned grains of gunpowder on epithelial tissue. Tattooing has been recorded with distances as close as 1 cm and as far away as 100 cm. The density of the tattooing will be principally determined by the muzzle to skin distance but may also be affected by the length of the gun barrel and the type of gun powder used.

Close Range Wounds

Close range entrance wounds are usually characterized by the presence of a surface contaminant known as soot (**Fig. 4A** and **Fig. 4B**). Soot is the carbonaceous byproduct of combusted gunpowder and vaporized metals. It is generally associated with a range of fire of less than 10 cm, but has been reported in wounds inflicted at distances of up to 20–30 cm. The density of the soot will decrease as the muzzle-to-skin distance decreases. The concentration of soot is also influenced by the amount and type of gunpowder used, the gun's barrel length, the caliber of the ammunition and the type of weapon used. For close range wounds, the longer the barrel length, the denser the pattern at a given distance. At a close range of fire, the visibility of tattooing may be obscured by the presence of soot. Also, at a close range, the unburned grains of powder may be injected directly into the entrance wound.

The term 'powder burns' is one used to describe a thermal injury to the skin, associated exclusively with the rise of weapons (muzzle loaders and starter pistols). When black power ignites, it produces a flame and large amounts of white smoke. The flame sometimes catches the user's clothing on fire, resulting in powder burns. Black powder is not used in any commercially available ammunition today. The expression powder burns should *not* be used to describe the carbonaceous material deposited on the skin with entrance wounds inflicted by commercially available cartridges.

Contact wounds

Contact wounds are wounds which occur when the barrel of the gun is in actual contact with the clothing or skin of the victim. There are two types of contact wounds: loose contact and tight. Tight contact wounds occur when the muzzle is pressed tightly against the skin. All material that is discharged from

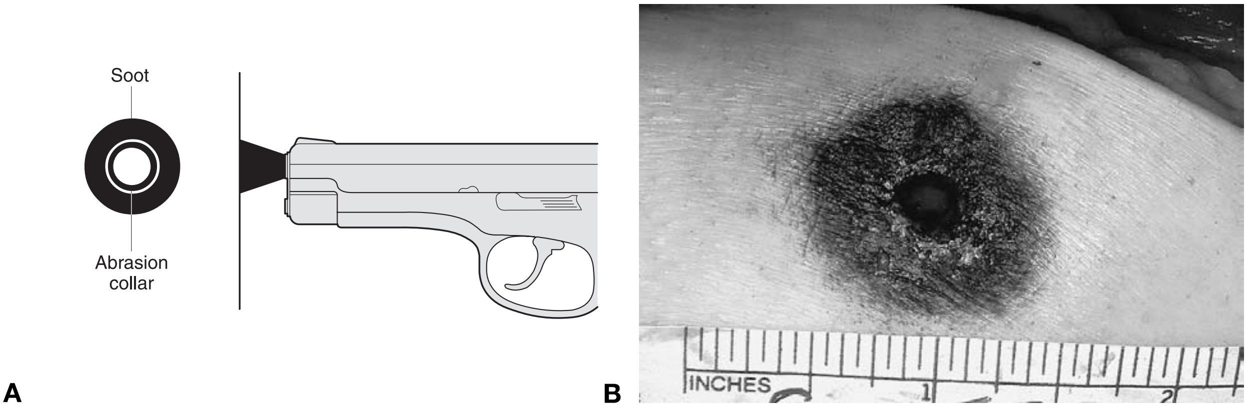


Figure 4A, B Close range gunshot wounds are characterized by the presence of soot. Soot is a carbonaceous material which will deposit on skin or clothing at distances generally less than 10 cm.

the barrel, including soot, gases, incompletely burned gunpowder, metal fragments, and the projectile are injected into the wound. In a loose contact wound, the contact between the skin and the muzzle is incomplete, and soot and other residues will be distributed along the surface of the epithelium.

When a tight contact wound occurs in an area of thin tissue or tissue overlying bone, the hot gases of combustion will cause the skin to burn, expand and rip (**Fig. 5**). The tears of the skin will appear triangular or stellate in configuration (Fig. 5). These stellate tears are commonly misinterpreted as exit wounds when the physician is basing that opinion on the size of the wound alone. The stellate wounds resulting from contact with the barrel will always be associated with the presence of seared skin and soot (**Fig. 6**). The stellate tears of exit wounds will lack soot and seared skin. The charred or seared skin of contact wounds will have the microscopic characteristics of thermally damaged skin. The injection of gases into the skin may also cause the skin to be forcibly compressed against the barrel of the gun and may leave a muzzle contusion or muzzle abrasion surrounding the wound (**Fig. 7A**, and **B**).

The size of gunshot wounds of entrance bears no reproducible relationship to the caliber of the projectile.

Exit Wounds

Exit wounds will assume a variety of shapes and configurations and are *not* consistently larger than their corresponding entrance wounds. The exit wound size is dictated primarily by three variables: the amount of energy possessed by the bullet as it exits the skin, the bullet size and configuration, and the amount of energy transferred to underlying tissue,

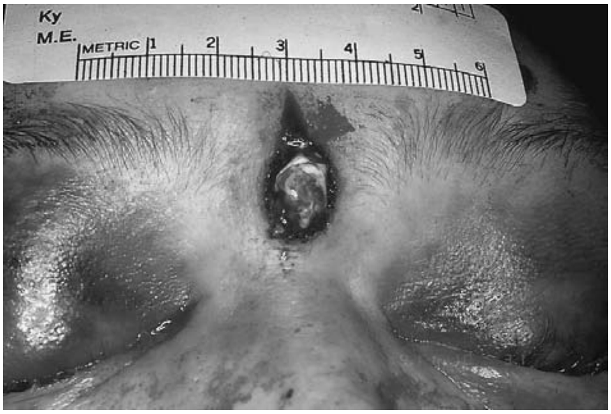


Figure 5 (see color plate 13) The contact wound will exhibit triangular shaped tears of the skin. These stellate tears are the result of injection of hot gases beneath the skin. These gases will cause the skin to rip and tear in this characteristic fashion.

Figure 6 Contact wounds will also exhibit seared wound margins as well as stellate-shaped tears. This is a contact wound from a 38 caliber revolver.

i.e. bone fragments. Exit wounds usually have irregular margins and will lack the hallmarks of entrance wounds, abrasion collars, soot, and tattooing (**Fig. 8A** and **B**). If the skin of the victim is pressed against or supported by a firm object, as the projectile exits, the wound may exhibit an asymmetric area of abrasion.

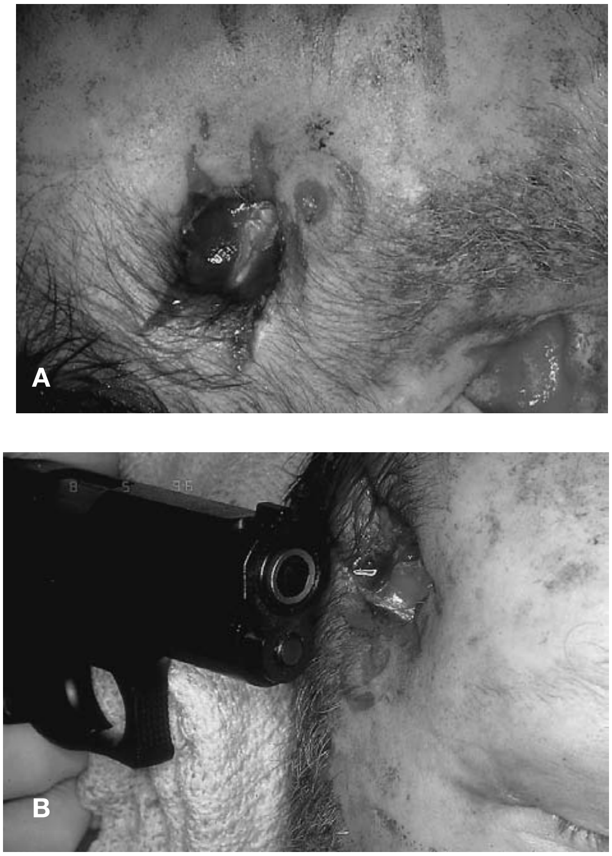
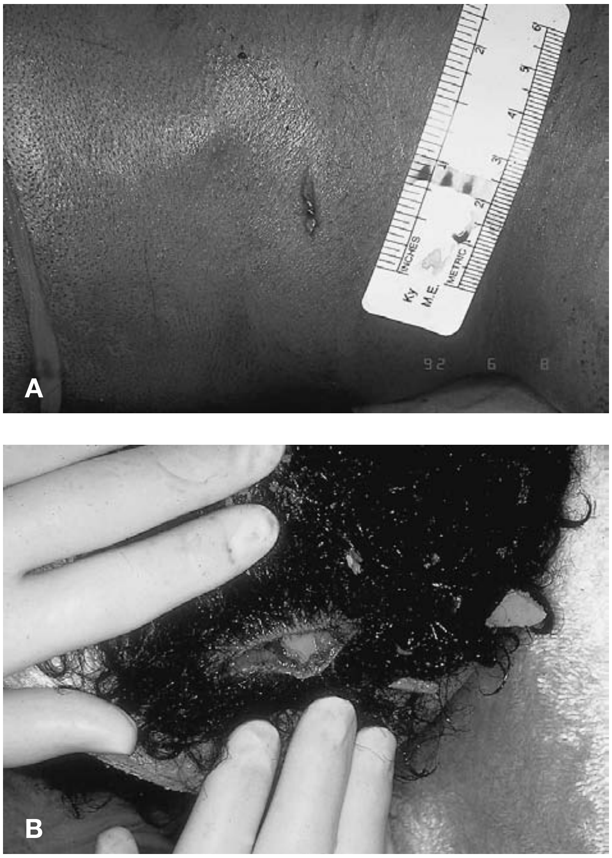


Figure 7A, B The injection of gases into the skin will cause the skin to expand and make forceful contact with the barrel of the gun (**A**). If there is sufficient gas, this contact will leave a 'muzzle contusion' or 'muzzle abrasion' around the wound. This pattern injury mirrors the end of the barrel (**B**).



Figures 8A, B Gunshot wounds of exit generally have irregular wound margins and will lack the physical evidence associated with entrance wound including abrasion collars, soot, seared skin and 'tattooing'. Exit wounds associated with handgun ammunition are not consistently larger than the associated exit wound. These slit-like exit wounds have irregular margins.

This area of abrasion has been described as a 'false' abrasion collar and is termed a 'shored' exit wound (**Fig. 9**). This false abrasion collar results when the epithelium is forced outward and makes contact or is slapped against the supporting structure. Examples of supporting structures are chair backs, floors, walls, or tight clothing. The shored exit wound may also be called a 'supported' exit wound.

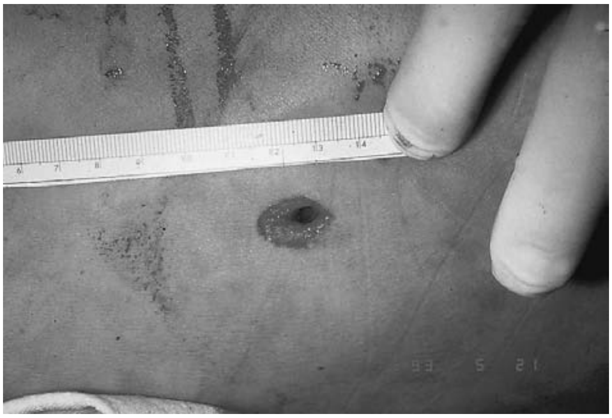


Figure 9 A 'shored' exit wound has the appearance of a 'false' abrasion collar. This false abrasion collar results when epithelium is forced outward and makes contact or is slapped against a supporting structure, i.e. floor, wall or furniture. A short exit may also be referred as a supported exit wound.

Evidence Collection

Any tissue excised from a gunshot wound margin should be submitted for microscopic evaluation. This microscopic examination will reveal the presence of carbonaceous materials, compressed and distorted epithelial cells and thermally induced collagenous changes in the dermal layer. This information may assist the practitioner in determining the range of fire or entrance versus exit when the physical exam is indeterminate. It is imperative that treating physicians recognize the importance of preserving evidence in the gunshot wound victim. It has been recognized that victims of gunshot wounds should undergo some degree of forensic evaluation prior to surgical intervention. It is also necessary that healthcare practitioners recognize, preserve and collect short-lived evidence and not contaminate or destroy evidence while rendering care.

The clothing of a gunshot wound victim may yield valuable information about the range of fire and will aid physicians in their efforts to distinguish entrance from exit wounds (**Fig. 10**). Fibers of clothing will deform in the direction of the passing projectile (**Fig. 11**). Gunpowder residues and carbonaceous soot will deposit on clothing, just as they do on skin. Some of these residues are not visible to the naked eye and require standard forensic laboratory staining techniques to detect the presence of lead and

Figure 10 Carbonaceous material as well as gunshot residue may be deposited on victims' clothing if the handgun is discharged within range of fire of less then 1 m. Some gunshot residue from nitrates and vaporized lead will be invisible to the naked eye. All clothing should be collected and placed in separate paper containers for evaluation by the forensic laboratory.

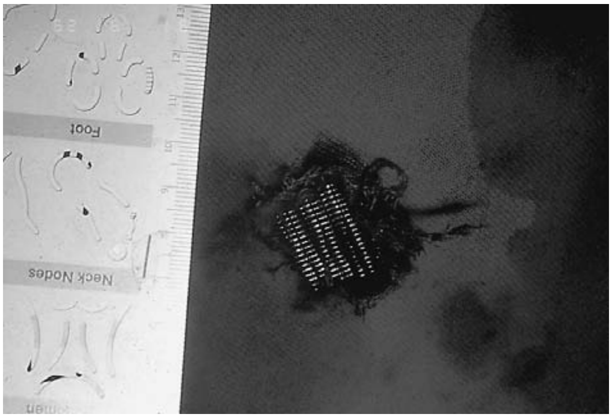


Figure 11 Close examination of fibers may reveal the direction of the passing projectile.

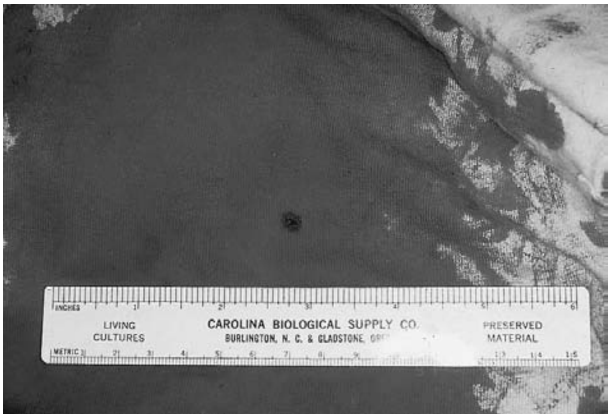


Figure 13 A lead or well-lubricated projectile may leave a ring of lead or lubricant on clothing as it passes through. This deposition of lubricant or lead is referred to as 'bullet wipe'.

nitrates (**Fig. 12**). Some bullets, in particular lead bullets, may deposit a thin layer of lead residue on clothing as they penetrate; this residue is termed 'bullet wipe' (**Fig. 13**).

When collecting articles of clothing from a gunshot wound victim, each article must be placed in a separate paper bag in order to avoid crosscontamination.

Some jurisdictions may elect to perform a gunshot residue test to determine the presence of invisible residues on a suspects' skin. Residues from the primer, including barium nitrate, antimony sulfide, and lead peroxide may be deposited on the hands of the individual who fired the weapon. The two methods used to evaluate for residues are flameless atomic absorption spectral photometry (FAAS), and scanning microscope–energy dispersive X-ray spectrometry (SEM-EDX). For analysis by FAAS, a specimen is collected by swabbing the ventral and dorsal surfaces of the hand with a 5% nitric acid solution. For analysis by SEM-EDX, tape is systematically pressed against the skin and packaged. A second method involves placement of tape on the hands and examining the material under a scanning electron microscope. The specificity and sensitivity of these tests have been questioned, as residues may be spread about the crime scene. These residues may be spread by secondary contact with a weapon or furniture items, and have been known to result in false positive tests. Controlled studies involving the transfer of gunshot residues from individuals who fired a weapon to suspects being handcuffed, have been documented. If a gunshot residue test is to be performed on a victim, it should be done within the first hour of the weapon's discharge to increase the sensitivity of the test.

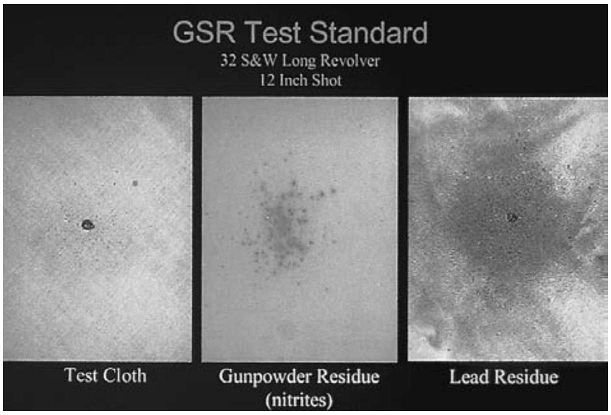


Figure 12 (see color plate 14) The forensic laboratory can use tests to detect the presence of vaporized lead and nitrates. These tests will assist the forensic investigator in determining the range of fire.

The bullet, the bullet jacket and the cartridge are invaluable pieces of forensic evidence. The forensic firearms examiner may use these items to identify or eliminate a weapon as the one which was fired. When these items are collected from the living patient, they should be packaged in a breathable container, a container that allows a free exchange of outside air. If such an item were placed in a closed plastic container, mold may form and evidence degrade.

When handling evidence, all metal-to-metal contact should be avoided. The tips of hemostats should be wrapped in sterile gauze and all evidence placed immediately in containers for transport (skip the metal basin step). Metal-to-metal contact may destroy or obliterate critical microscopic marks.

Estimation of Bullet Caliber Using Radiographs

Radiographs taken to assist the treating physician in locating retained projectiles may also be of evidentiary value. Radiographs will assist the physician in determining the direction from which projectiles were

fired as well as their simple number. Practitioners should not render opinions as to the caliber of a specific projectile, based upon radiographic imaging alone, because only radiographs taken exactly 72 in (183 cm) from the projectile will reveal the projectile's appropriate size. The bullet size on the radiographic film will increase as the distance between the film and X-ray source decreases.

Conclusion

If investigators and health practitioners receive adequate forensic training, the distinction of entrance from exit wounds will not be difficult. When practitioners are unsure, they should refrain from placing statements regarding entrance and exit in the medical record. The physician and patient would be better served if the wound were measured, photographed and accurately described, using appropriate forensic terminology and sample tissues sent to the forensic pathologist for microscopic analysis. Of course, the primary concern of all practitioners caring for survivors of gunshot wounds is the health and well-being of their patient; however, a simple awareness of and attention to the forensic issues involved in gunshot wound examination and management can profoundly affect the medicolegal/justice aspects of the victim's experience.

See also: **Firearms**: Weapons, Ammunitions and Penetration; Range; Residues; Humane Killing Tools; Laboratory Analysis.

Further Reading

Carmona R and Prince K (1989) Trauma and forensic medicine. *Journal of Trauma* **29**:1222.

Collins KA and Laintz PE (1994) Interpretation of fatal, multiple and exiting gunshot wounds by trauma specialists. *Journal of Forensic Science* **39**:94.

DiMaio VJM (1999) *Gunshot Wounds*, 2nd edn. Boca Raton: CRC Press.

Dixon DS (1988) Gunshot wounds: forensic implications in a surgical practice. In: Ordog GJ (ed.) *Management of Gunshot Wounds*. New York: Elsevier.

Eckert W (1992) The development of forensic medicine in the United Kingdom from the eighteenth century. *American Journal of Forensic Medical Pathology* **13**:124.

Fackler ML (1988) Wound ballistics. A review of common misconceptions. *Journal of the American Medical Association* **259**:2730.

Fackler ML and Riddick L (1996) Clinicians' inadequate descriptions of gunshot wounds obstruct justice – clinical journals refuse to expose the problem. *Proceedings of the American Academy of Forensic Science* **2**:150.

Gliamas DM *et al* (1995) Officers, their weapons and their hands: an empirical study of GSR on the hands of nonshooting police officers. *Journal of Forensic Science* **40**:1086.

Godley DR and Smith TK (1997) Medicolegal aspects of gunshot wounds. *Journal of Trauma* **17**:866.

Marlowe AL and Smock WS (1996) The forensic evaluation of gunshot wounds by prehospital personnel. *Proceedings of the American Academy of Forensic Science* **2**:149.

Randall T (1993) Clinicians' forensic interpretations of fatal gunshot wounds often miss the mark. *Journal of the American Medical Association* **269**:2058.

Smialek JE (1983) Forensic medicine in the emergency department. *Emergency Medical Clinics of North America* **1**:1685.

Smock WS, Nichols GR and Fuller PM (1993) Development and implementation of the first clinical forensic medicine training program. *Journal of Forensic Science* **38**:835.

Smock WS (1994) Development of a clinical forensic medicine curriculum for emergency physicians in the U.S.A. *Journal of Clinical Forensic Medicine* **1**:27.

Recognition of Pattern Injuries in Domestic Violence Victims

W S Smock, University of Louisville School of Medicine, Louisville, KY, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0751

Introduction

Victims of domestic violence present regularly to emergency departments. Several studies have indicated that 22–33% of all patients who present to emergency departments are victims of domestic violence. This puts emergency physicians and nurses in a unique position to screen for victims of this societal plague. The literature suggests, however, that only a small percentage of victims of domestic violence are recognized as such in the emergency department. These studies document detection rates by emergency staff as low as 5.6%. One study that evaluated emergency department visits by victims of domestic violence found that 43% of women with acute trauma presented six or more times before they were

recognized and identified as victims of abuse; nearly half these patients presented no less than 12 times. These findings, though they may shock us, should perhaps not surprise us, given that fewer than 50% of the medical schools in the United States include courses on domestic violence in their curriculum. This deficiency is also present in residency training programs. A study released in 1989 reported that only 8% of practicing physicians indicate they have received adequate training to deal with issues related to and victims of domestic violence.

Identification of Pattern Injuries

A pattern injury is one which has a tell-tale marker for the tool which inflicted it. The list of 'tools' is infinite. Commonly used ones are: hands, belts, baseball bats, kitchen utensils and curling irons. Each of these tools leaves a skin imprint which reflects its shape and is, therefore, unique to it. Pattern injuries are consistently reproducible. They can be classified into three major categories, according to their source: blunt force, sharp force and thermal.

A knowledge of pattern injuries and the precise documentation as to the anatomic location of each injury will assist the physician and law enforcement officer in determining what implement, tool or weapon was responsible for producing a particular wound. Knowing which tool or weapon was used will help the investigator to determine if the injury is consistent or inconsistent with the history given by the patient, a caregiver or an accused assailant. In addition to written documentation, photographic documentation of injuries is desired for placement in the patient's medical record.

Blunt Force Pattern Injuries

Contusions

The most common blunt force injury is the contusion or bruise. Other manifestations of blunt force/trauma to the skin are the abrasion and the laceration. With sufficient force, the offending weapon, whose shape or configuration is unique, will stamp a mirror image of itself on the skin. Examples commonly seen in victims of assault, domestic violence and child abuse include: slap marks from the hand with digits delineated (**Fig. 1**), looped or flat contusions from belts or cords (**Fig. 2**), contusions from fingertip pressure (**Fig. 3**), scratches from fingernails (**Fig. 4**), parallel contusions from contact with a linear object, like a baseball bat (**Fig. 5**), contusions from the heels and soles of shoes (**Fig. 6**), and semicircular contusions or contused abrasions from bites (**Fig. 7**).

Figure 1 The hand is the most common 'weapon' of domestic violence. The slap mark present on this patient's cheek shows areas of central clearing from impact with the extended fingers.

A blow from a linear object leaves a contusion characterized by a set of parallel lines separated by an area of central clearing. The blood directly underneath the striking object is forcibly displaced to the sides, which accounts for the pattern's appearance (Fig. 5 and **Fig. 8**).

Although some injuries are not remarkable or important in terms of physical treatment, emergency physicians should be suspicious upon encountering them. When physicians observe circular contusions or

Figure 2 (see color plate 15) Blows from a belt can exhibit several different patterns. If the belt impacts on edge, a linear contusion will be produced. If the belt impacts more on its side, a wider contusion may be imprinted. Woven belts will leave a mirror image of the weave imprinted on the skin.

Figure 3 Pressure from fingertips will leave small semicircular contusions, measuring 1.0–1.5 cm in diameter.

Figure 5 Impact with a baseball bat will result in an area of central clearing surrounded by two parallel lines. The blood directly beneath the area of impact is forced upward and outward around a linear object, resulting in rupture of the vessel outside the area of impact.

linear contusions, they should be suspicious of abuse or assault. Circular contusions 1.0–1.5 cm in diameter are consistent with fingertip pressure and grab marks (Fig. 3). An anatomical location where fingertip pressure contusions are often seen and easily overlooked is the medial aspect of the upper arm (**Fig. 9**). When a physician suspects a linear contusion to be a belt mark, a very careful examination of the wound may reveal evidence which will aid law enforcement officers in searching a domicile to find the offending belt in question.

The most common history given by a victim of domestic violence is: 'I fell'. When attempting to determine if the history is consistent with the physical findings, one must evaluate where the injuries are located, and where they are absent. When an individual falls, one expects to find blunt force trauma present at bony prominence, including: the zygomatic arch, elbows, extensor services or the forearm and knees (**Fig. 10**). When soft tissue trauma is seen in other anatomical locations, particularly breasts (**Fig. 11**), chest, back, abdomen and neck, one must have a high index of suspicion for inflicted injury.

The pattern contusion which may be the most specific to and sensitive for the identification of the individual who inflicted it, is the bite mark. The bite mark may present as a pattern contusion, a pattern abrasion or a combination of the two (Fig. 7). Bite marks vary greatly in the quality of their identifiable features, dictated by the anatomical location of the bite and the motion of the teeth relative to the skin. Some bite marks may not be identified as a bite, and present as a nonspecific abrasion or contusion (**Fig. 12**). When a bite is identified in an acute setting and the perpetrator is 'unknown', care must be taken not to destroy potential evidence. Evidence, including buccal cells with their inherent DNA and saliva cells

Figure 4 Contact from fingernails will result in a pattern abrasion which may be associated with up to four parallel linear abrasions.

Figure 6 A kick or stomp from the sole of a shoe may give rise to a pattern contusion, which mirrors the pattern of the sole. Such contusions must be documented early in the course of patient care, prior to the fading or resolution of the pattern.

Figure 7 Bite marks may have an array of appearances including: semicircular, arched contusions, abrasions, and imprints of individual teeth characteristics.

Figure 9 The medial aspect of the upper arm demonstrates a 'grab mark'. Pressure from fingers and fingertips will give rise to fingertip contusions as well as pattern contusions, which outline the perpetrator's fingers.

with their inherent ABO blood group protein antigens could be easily washed away if a bite were cleaned prior to the forensic evaluation. The first step in 'suspicious wound' management should be to swab the skin's surface with a sterile cotton swab, moistened with sterile saline or distilled water. This swabbing may succeed in collecting traces of the assailant's saliva and epithelial cells. Blood group antigens degrade rapidly. This evidence is short-lived and should, therefore, be managed with this fact in mind. About 80% of the population will secrete their ABO blood group protein antigens in their saliva. This, plus the collection of epithelial cells, with DNA, will assist law enforcement agencies in the identification of or exclusion of certain suspects. Ideally, a forensic odontologist should be consulted in an effort to document the wound in a manner which permits the comparison of the teeth marks with the teeth of the suspect at a later date.

Utilization of alternative light sources such as ultraviolet or infrared may reveal a pattern contusion within or deep under the epithelium, which is not visible to the naked eye. This is most helpful in patients who are darkly pigmented (**Fig. 13**). Such sources are routinely used by forensic odontologists on faint, old or difficult bite marks. Successful identifications have been made on bite marks at six months postinjury.

The dating or aging of a contusion or bruise has been the focus of much debate within the forensic community. The development of a contusion is based on a number of variables: amount of force applied to

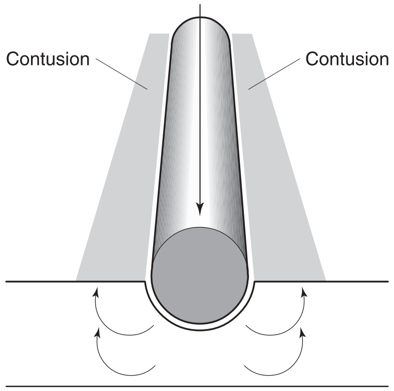


Figure 8 A diagrammatic representation of the pathophysiology associated with blunt force impact on a rounded linear object. The blood directly beneath the rounded linear object is forced up to the side and the increased pressure under vessels causes the vessels to rupture. The area of central clearing surrounded by two parallel lines is pathognomonic for impact with a rounded linear object.

Figure 10 Injuries that are associated with falls would include contusions and abrasions to the bony surfaces of the body, including: the zygomatic arch, extensor surfaces of the arms, elbows, knees and shins.

Figure 11 The presence of unexplained contusions to the breast is strongly correlated with domestic or partner violence. This breast displays a several-day-old contusion from the tissue being squeezed during an assault.

the skin, the vascularity of the tissue, the fragility of blood vessels, the tissue density and the amount of blood that escapes into the surrounding tissue. As a result, there is *no* reproducible standard for the dating of a contusion based on its color. This author does not recommend the dating of a contusion based on the color to the naked eye. At present there are projects ongoing to evaluate the dating of contusions, based on degradation of the heme molecule. Currently these studies are experimental but are showing promise.

Pattern abrasions

An abrasion is a rubbing or scraping away of the superficial layers of the epidermis. Examples of pattern abrasions include fingernail scratches (Fig. 4), bite marks, imprints of carpet fabric (**Fig. 14**), and ligature marks around the wrist or neck. The presence of such pattern injuries, unremarkable from the standpoint of treatment, may provide invaluable information from a forensic and injury reconstruction perspective.

Figure 12 This is an 'atypical' bite mark from the victim pulling away from the teeth of the perpetrator. This resulted in linear or curvilinear abrasions.

Figure 13 The use of infrared light sources can assist law enforcement in determining the presence of blood beneath the skin in darkly pigmented patients. This patient exhibits contusions which were invisible to the naked eye.

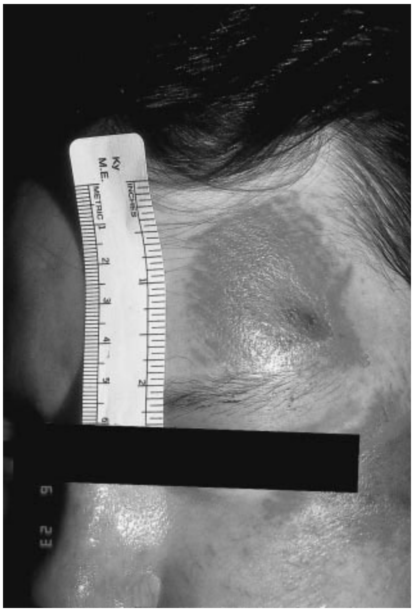
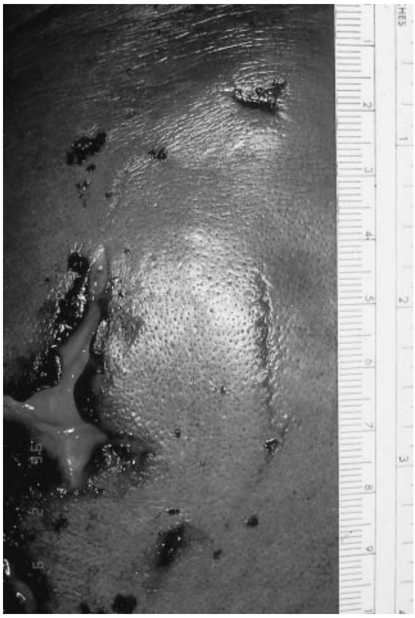


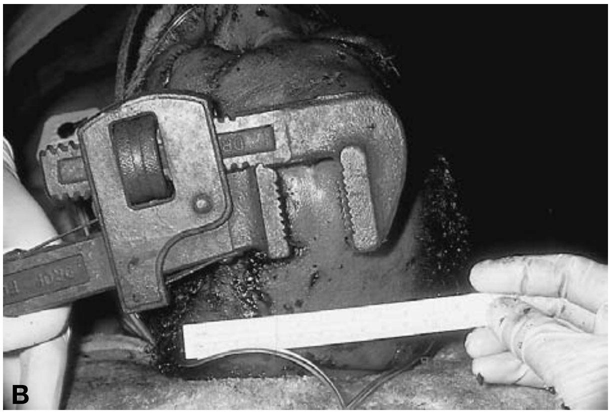
Figure 14 The pattern abrasion present on the patient's left forehead was the result of impact with a carpeted floor. The linear characteristics of the carpet's weave are imprinted on the patient's skin.

Pattern lacerations

A laceration is defined as a 'tear produced by blunt force trauma'. The term laceration is often misused by medical professionals to describe an incised wound. An incised wound results from a sharp-edged implement, i.e. scalpel, knife or glass shard being drawn across the skin. To distinguish a laceration from an incised wound, one needs simply to look for the laceration's characteristic abraded or crushed skin edges, and unique tissue-bridging. In contrast, the edges of an incised wound will be sharply demarcated. Pattern lacerations are the least common type of pattern injury. An example of a pattern laceration is shown in **Fig. 15**; the wave-like wound imprint was produced by the head of a pipe wrench.



A



B

Figure 15 A,B This pattern laceration on the patient's scalp has wavelike characteristics which were imprinted from the teeth of a pipe wrench.

Sharp Force Pattern Injuries

Sharp force injuries are of two types: either incised or stabbed. The incised wound is longer than it is deep. The stab wound is a puncture wound that is deeper than it is wide. The wound margins of sharp force injuries are clean and lack the abraded edges of injuries from blunt force trauma.

A certain amount of forensic information can be gathered from the examination of a stab wound. Characteristics of a knife blade, single or double edged, can be determined from visual inspection (**Fig. 16**). Additional characteristics such as serrated versus sharp, can be detected if the blade was drawn across the skin during its insertion or withdrawal (**Fig. 17**). It is important to note, however, that serrated blades do not always leave characteristic marks.

Thermal Pattern Injuries

A thermal pattern injury is an injury whose offending force is heated and whose physical appearance belies the heat source. Thermal injuries of an inflicted nature are commonly seen in the elderly and in children. The detailed history of a thermal incident should include how the burn occurred and should also elicit information about the position of the patient relative to the thermal source. This information is critical in order to render an opinion as to whether the injury was of an inflicted or accidental nature. Common inflicted thermal pattern injuries include: flat iron burns (**Fig. 18**), curling iron burns (**Fig. 19**), splash burns (**Fig. 20**) and immersion burns (**Fig. 21**).

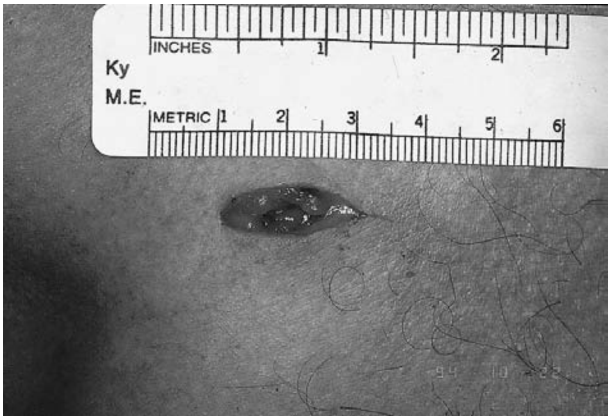


Figure 16 Examination of a stab wound will reveal the characteristics of the inflicting knife. This wound demonstrates a sharp and a dull edge. The sharp edge is associated with the sharp edge of the knife blade.

Figure 17 Examination of this wound reveals small punctate abrasions near the sharp edge. These abrasions are the result of contact with the serrated portions of the blade. A serrated blade may however inflict no abrasions if the blade is inserted and withdrawn in a vertical fashion.

Figure 18 (see color plate 16) Contact with the flat portion of a hot iron results in a unique thermal injury. The areas of sparing are from the steam holes.

Figure 19 Contact with a linear curling iron will result in a mirrored linear thermal injury.

Figure 20 'Splash burns' are the result of splashed water droplets. Areas of satellite lesions are expected from splash burns.

Immersion or dipping burns are ones characterized by a sharp or clear line of demarcation between burned and unburned tissue. Splash burns, in contrast, are characterized by an irregular or undulating line over isolated areas of thermal injury, usually round or oval in shape, caused by droplets of hot liquid. The severity of the thermal or scald injury is dependent on the length of time that the skin is in contact with the offending source and the temperature of the source itself. In the case of water, from the tap or faucet, it would take 1 s to cause a full thickness burn at 150° (65.5°C), and 180 s to elicit an injury of comparable severity at 120° (49°C) (**Fig. 22**). The measurement by law enforcement of the water temperature of a households or institutions should be a standard component of any investigation involving in a scald injury of a child, a developmentally delayed patient or an elderly individual.

Figure 21 When a child's foot is immersed in a hot liquid, a line of demarcation results. This patient had symmetrical burns on her right and left feet.

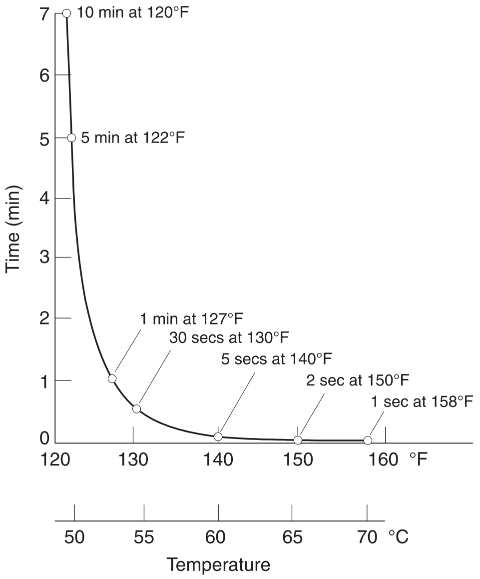


Figure 22 The factors associated with thermal injury are the temperature of the offending medium and the duration of contact.

Conclusion

Familiarity with pattern injuries will assist physicians, nurses and law enforcement investigators with the recognition of inflicted injuries in victims of child abuse and domestic violence. The ability to differentiate between fact and fiction will benefit the criminal justice system in its effort to protect the innocent and prosecute the guilty.

See also: **Causes of Death**: Blunt Injury; Sharp Injury; Burns and Scalds. **Clinical Forensic Medicine**: Child Abuse. **Odontology**.

Further Reading

Council on Scientific Affairs: American Medical Association (1992) Violence against women, relevance for medical practitioners. *Journal of the American Medical Association* 267:3184.

Delahunta EA: (1995) Hidden trauma: the most misdiagnoses of domestic violence. *American Journal of Emergency Medicine* 13:74.

Giardino AP, Christian CW and Giardino ER (1997) *A Practice Guide to the Evaluation of Child Physical Abuse and Neglect*. Thousand Oaks, CA: Sage.

Gremillion TH and Evins G (1994) Why don't doctors identify and refer victims of domestic violence? *North Carolina Medical Journal* 55:428.

Grunfield AF *et al.* (1994) Detecting domestic violence against women in the emergency department: a nursing triage model. *Journal of Emergency Nursing* 20:271.

Katcher ML (1981) Scald burns from hot tap water. *Journal of the American Medical Association* 24:1219.

Morrison LJ (1988) The battering syndrome: a poor record of detection in the emergency department. *Journal of Emergency Medicine* 6:521.

Reese RM (1994) *Child Abuse: Medical Diagnosis and Management*. Malvern, PA: Lea & Febiger.

Smock WS (1997) Clinical forensic medicine. In: Rosen (ed.) *Emergency Medicine*, 4th edn.

Spitz WU (1993) *Medicolegal Investigation of Death*. Springfield: Charles C. Thomas.

Tilden VP and Shepherd P (1987) Increasing the rate of identification of battered women in an emergency department: use of a nursing protocol. *Research Nursing Health* 10:209.

Self-inflicted Injury

S Pollak, University of Freiburg/Br, Freiburg, Germany
P J Saukko, University of Turku, Turku, Finland

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0439

Introduction

The physical examination of persons with injuries that have been self-inflicted deliberately is part of the routine work in clinical forensic medicine. The recognition of self-harm is of criminalistic importance when differentiating between assaults, accidents, suicidal acts and other kinds of self-destructive behavior; the main categories are listed in **Table 1**. In cases of

Table 1 Main categories of self-induced bodily harm

1. Simulation of a criminal offense:
   (a) False rape allegation
   (b) Feigned robbery or alleged attack for other fabricated reasons
2. Self-mutilation for the purpose of insurance fraud
3. Voluntary self-mutilation and/or malingering among soldiers and prisoners
4. Dermal artifacts, self-mutilation and malingering in patients with personality disorders
5. Suicidal gestures, (attempted) suicide

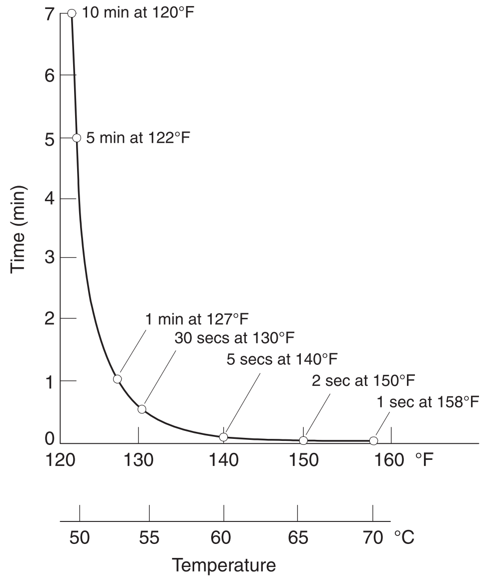


Figure 22 The factors associated with thermal injury are the temperature of the offending medium and the duration of contact.

Conclusion

Familiarity with pattern injuries will assist physicians, nurses and law enforcement investigators with the recognition of inflicted injuries in victims of child abuse and domestic violence. The ability to differentiate between fact and fiction will benefit the criminal justice system in its effort to protect the innocent and prosecute the guilty.

See also: **Causes of Death**: Blunt Injury; Sharp Injury; Burns and Scalds. **Clinical Forensic Medicine**: Child Abuse. **Odontology**.

Further Reading

Council on Scientific Affairs: American Medical Association (1992) Violence against women, relevance for medical practitioners. *Journal of the American Medical Association* 267:3184.

Delahunta EA: (1995) Hidden trauma: the most misdiagnoses of domestic violence. *American Journal of Emergency Medicine* 13:74.

Giardino AP, Christian CW and Giardino ER (1997) *A Practice Guide to the Evaluation of Child Physical Abuse and Neglect*. Thousand Oaks, CA: Sage.

Gremillion TH and Evins G (1994) Why don't doctors identify and refer victims of domestic violence? *North Carolina Medical Journal* 55:428.

Grunfield AF *et al*. (1994) Detecting domestic violence against women in the emergency department: a nursing triage model. *Journal of Emergency Nursing* 20:271.

Katcher ML (1981) Scald burns from hot tap water. *Journal of the American Medical Association* 24:1219.

Morrison LJ (1988) The battering syndrome: a poor record of detection in the emergency department. *Journal of Emergency Medicine* 6:521.

Reese RM (1994) *Child Abuse: Medical Diagnosis and Management*. Malvern, PA: Lea & Febiger.

Smock WS (1997) Clinical forensic medicine. In: Rosen (ed.) *Emergency Medicine*, 4th edn.

Spitz WU (1993) *Medicolegal Investigation of Death*. Springfield: Charles C. Thomas.

Tilden VP and Shepherd P (1987) Increasing the rate of identification of battered women in an emergency department: use of a nursing protocol. *Research Nursing Health* 10:209.

Self-inflicted Injury

S Pollak, University of Freiburg/Br, Freiburg, Germany
P J Saukko, University of Turku, Turku, Finland

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0439

Introduction

The physical examination of persons with injuries that have been self-inflicted deliberately is part of the routine work in clinical forensic medicine. The recognition of self-harm is of criminalistic importance when differentiating between assaults, accidents, suicidal acts and other kinds of self-destructive behavior; the main categories are listed in **Table 1**. In cases of

Table 1 Main categories of self-induced bodily harm

1. Simulation of a criminal offense:
   - (a) False rape allegation
   - (b) Feigned robbery or alleged attack for other fabricated reasons
2. Self-mutilation for the purpose of insurance fraud
3. Voluntary self-mutilation and/or malingering among soldiers and prisoners
4. Dermal artifacts, self-mutilation and malingering in patients with personality disorders
5. Suicidal gestures, (attempted) suicide

simulated offenses, early diagnosis prevents superfluous and inevitably unsuccessful investigations by the police and mystified worries by the public. In the context of this article, the chief stress is put on the medicolegal and criminalistic aspects of physical traumatization by the victim himself. Completed (fatal) suicides, nonaccidental poisonings, and self-induced bodily harm in psychiatric patients are dealt with in other articles.

Simulation of Criminal Offenses

This group comprises self-inflicted injuries of individuals who claim to have been the victim of an assault. The bodily damage, therefore, is used as alleged proof of a fictitious offense. The dramatic story told by the informant is often in obvious contrast to a uniform wound pattern (**Fig. 1**) and a poor severity of the injuries so that the whole picture does not suggest a real struggle with the dynamics of a fight.

According to the literature, up to 20% of the sexual offenses reported to the police are fictitious. In false rape allegations the informants frequently injure themselves in order to give support to their story (**Fig. 2**). Women and girls who falsely claim to have been raped often live in a problematic situation or in conflict with their partners. The primary intention of the alleged victims is mostly to derive attention, or care and affection, but there are also some informants who accuse a definite person of having raped them for motives of hate or revenge.

Usually the self-inflicted injuries are caused with the help of pointed and/or cutting tools such as knives, razor blades, nail-scissors, and broken glass. The resulting wounds are of a trivial nature, mostly consisting of superficial cuts or linear abrasions. The characteristics of a 'classical' injury pattern are listed in **Table 2**. In typical cases there is a multitude of equally shallow lesions which are strikingly uniform in shape, often orientated in the same direction or in a criss-cross manner. The cuts avoid especially sensitive areas like the eyes, lips, nipples, and genitals. They are mainly located on the frontal aspect of the trunk, on the face, the neck, the arms and hands, and sometimes

Figure 1 Group of mostly parallel skin lesions inflicted on the right cheek with nail scissors (23-year-old woman, who falsely claimed to have been the victim of a sexual assault; actually she wanted to arouse the attention of her unfaithful husband).

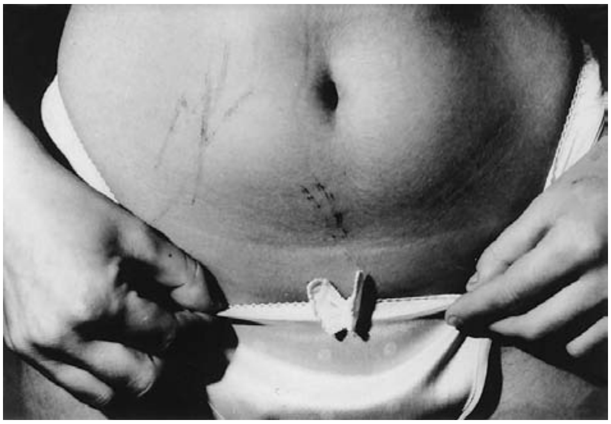


Figure 2 Self-inflicted injuries from a serrated kitchen knife on the lower abdomen of a young woman. She reported her ex-lover to the police in revenge for having been left and falsely claimed that he had raped her. The cuts are equally superficial and do not penetrate the skin in spite of their considerable length and the curved body surface.

Table 2 Typical features of self-inflicted injuries fabricated to simulate a criminal offence

1. Infliction either by sharp/pointed instruments or by fingernails
2. Equally shallow, nonpenetrating cuts or fingernail abrasions (sometimes each of considerable length)
3. Multitude of individual lesions
4. Uniform shape, linear or slightly curved course of the lesions
5. Grouped and/or parallel and/or criss-cross arrangement
6. Symmetry or preference of the nondominant side of the body (usually the left)
7. Location in easily reachable body regions
8. Omission of especially sensitive body regions
9. No damage of the clothes or inconsistent damage
10. Lack of defense injuries
11. Additional presence of scars from former self-injurious behavior

on the lower limbs. Nevertheless, self-inflicted injuries may also be found on the back of the body, as far as it is accessible for the alleged victim's hand, or if the injuries have been caused with the assistance of another person. Often the relevant garments are either undamaged or the damage does not correspond to the skin lesions.

Some women who claim a fictitious sexual assault inflict blunt skin injuries on themselves: for instance linear excoriations produced by scratching with the person's own fingernails or by rubbing the skin against a rough surface. In rare cases, atypical findings such as contusions or singular cuts have been observed. As there are a few clinical reports on self-cutting after rape, the forensic expert has to be aware of uncommon self-destructive behavior, even in real victims with a post-traumatic stress disorder.

Apart from false rape allegations, there are some other categories of fictitious criminal offenses. Thus, an assault may be simulated and made credible by self-inflicted injuries in order to divert attention from the person's own theft or embezzlement. The authors have also seen several male 'victims' who claimed to have been stabbed or cut by perpetrators because they wanted to arouse compassion and regain the affection of their wife or girlfriend (**Fig. 3**). In the last ten years an increasing number of informants cut swastikas, or other nazi symbols or words into their skin thus trying to gain sympathy as pretended victims of right-wing violence; if such incidents are covered in the media, they may be followed by an 'endemic' series of copy-cat acts.

Another possible reason for claiming a feigned assault is dissimulation of an attempted, but not completed suicide (e.g. in cases of nonfatal wrist-cutting). Injuries from autoerotic manipulation may also be concealed by concocting a story of criminal victimization. False allegation of kidnapping in combination with self-injury has been made by young girls and boys in order to excuse absence without leave. From the criminalistic point of view it has to be considered that perpetrators may inflict injuries on themselves in order to represent a homicide as permitted self-defense.

Self-mutilation for the Purpose of Insurance Fraud

In this category of self-destructive behavior, an accident is simulated in a fraudulent attempt to obtain compensation from an insurance company. Usually the deliberate self-inflicted damage results in mutilation, i.e. a substantial loss of a peripheral part of the body (mostly a finger or a hand). The instruments

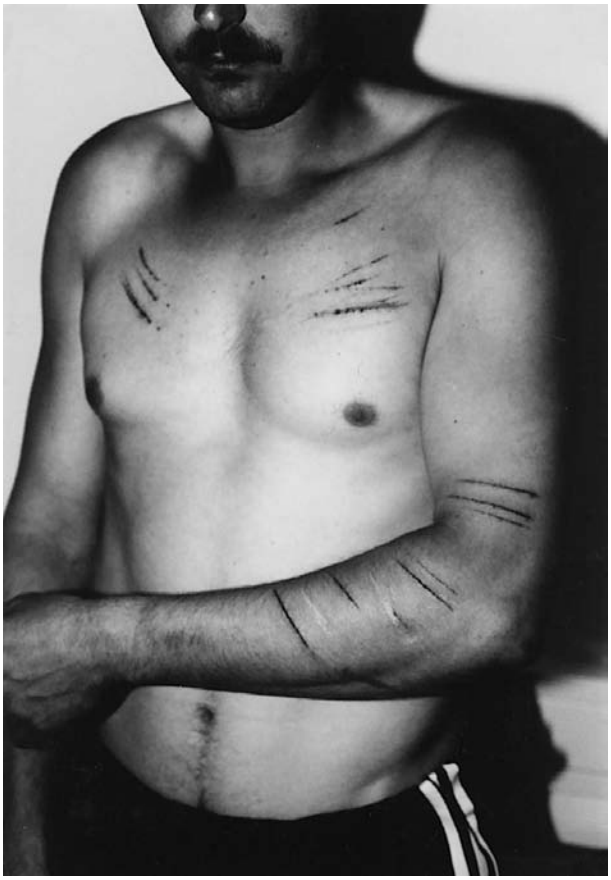


Figure 3 A 29-year-old right-handed man who inflicted cuts on his chest and left arm in order to gain the sympathy of his wife who had engaged in an extramarital affair with a lover. Note the symmetric arrangement of the chest wounds, their parallelism and uniformity. The whitish scars on the left forearm are due to former self-injurious episodes.

used for mutilation comprise sharp-edged tools such as axes, hatchets, meat cleavers and cutting machines, various kinds of motorized saws and, less frequently, blunt objects such as presses or conveyor belts.

Voluntarily inflicted injuries from axe blows are mostly claimed to have been caused by a misplaced stroke of the axe when chopping wood. In typical cases, the thumb or index finger is cut off completely in a right angle to its axis resulting in an isolated proximal finger amputation. In contrast, an unintentional severence of the index finger is usually characterized by accompanying injuries to the adjoining fingers. In authentic accidents, the amputation is often a distal and incomplete one, taking its course in an oblique angle to the finger's axis. The same is true for accidents with power saws or milling machines (**Fig. 4**).

It has been justly stressed that in finger injuries from heavy axe blows a complete amputation is to be

Figure 4 Left hand of a carpenter who injured himself accidentally when working with a milling machine. Note the oblique severence of the thumb and index finger. The distal phalanx of the middle finger was incompletely amputated and could be sewed on.

Figure 5 Self-mutilation for the purpose of insurance fraud. Radiograph of the left hand exhibiting a complete proximal amputation of the index finger at a right angle to its axis and without any injury to the adjoining fingers, allegedly caused by a misplaced stroke of the axe when chopping wood. (Photograph by courtesy of Professor Barz.)

expected only if the severed finger was lying on a solid base serving as a support. A proximal and complete amputation of the index finger without concomitant injuries of the neighboring fingers suggests a so-called 'execution position', which is extremely suspect of intentional self-mutilation (**Fig. 5**). In the last few years a considerable number of physicians – the majority being males over 40 years of age who worked in the surgical field – have been convicted of defrauding insurance companies by deliberately cutting off a finger, mostly the index finger of the nondominant left hand. Some reports have been published on cases in which members of medical professions applied a local anesthetic before injuring or mutilating themselves.

The proof of intentional self-infliction in cases of mutilation has been said to be one of the most difficult tasks in clinical forensic medicine. Apart from the medical aspects, some circumstantial indications may point to a possibly attempted insurance fraud: an insurance policy for inadequately high amounts taken out shortly before the injury occurs, a multiplicity of private accident insurance contracts, serious indebtedness of the policy holder, the absence of any witness, an inexplicable disappearance of the amputate, immediate tidying up of the scene, and removal of the biological traces. The presence of 'tentative cuts' adjacent to the amputation cut can be a positive morphological indication of deliberate self-infliction. It is recommended that the expert's investigation should first focus on the medical findings concerning the injury (photographs, surgical report, radiographs, physical examination) and on the course of the alleged accident as described and demonstrated by the mutilated victim. In addition, the external circumstances at the scene, the mechanical properties of the causative tool, the distribution of bloodstains and other available pieces of evidence must be taken into consideration. Some experts have also carried out simulating experiments to reconstruct the alleged

accident mechanism, which then turned out to be or not to be consistent with the injury in question.

Voluntary Self-mutilation and/or Malingering Among Prisoners and Soldiers

Autoaggressive behavior is a well-known problem in police custody and penal institutions. There may be different reasons for a prisoner to inflict injuries on himself: false allegations of having been physically maltreated by the police, the guard, or other inmates; the wish to be transferred to a hospital with less severe surveillance and increased chances of escape; as a correlate of a prisoner's low tolerance to stress. Common methods of self-harm under detention are cutting oneself with sharp instruments (for instance a piece of glass, a razor blade, or a sheet iron), swallowing foreign bodies, and certain forms of malingering (voluntary provocation, aggravation, and protraction of disease by artificial means). Self-damage in prison covers a wide continuum ranging from amateurish tattooing and infliction of other skin lesions to life-threatening suicide attempts.

During war and sometimes even in peace time soldiers may fabricate 'accidental' injuries in order to avoid duty on the frontline or in the armed forces generally. The deliberately induced damage is intended to make the soldier unfit for military service (for example because of traumatic amputation by shooting or cutting off fingers). Another kind of evading service aims at the pretence of sickness. By feigning a medical or psychiatric illness the malingerer tries to get hospitalized or dismissed. Malingering comprises both the simulation of a nonexisting illness and the exaggeration/prolongation of an existing disease.

Artifacts in Patients with Psychic Disturbances or Mental Diseases

In neurotic patients, the main target of bodily harm is the skin which may exhibit a great variety of lesions from scratching, pinching, squeezing, rubbing, and biting. The injuries are typically produced by the person's own fingernails, pointed/edged instruments, or rough surfaces. If the skin damage extends to the corium it will heal with scarring, and occasionally, pigmentation. The majority of neurotic excoriations are irregular in shape and located in easily reachable body regions, such as the chest and the lateral aspects of the face, arms, and thighs, often with preference of the nondominant side. The coexistence of fresh lesions and different stages of wound healing, including scars, hints at repeated episodes of self-injurious behavior (**Fig. 6**).

Figure 6 Left forearm of a 21-year-old depressive patient with a schizoid personality disorder and multiple episodes of self-cutting. Note the different wound age of scars and incisions.

If patients pretend to be unaware of the real origin of their skin artifacts, this phenomenon is called factitial dermatitis. The mysterious disease is intended to draw attention to the emotional suffering of the patient whose personality structure is characterized by strong intrapsychic tension, inhibited aggression, depressiveness, and low frustration tolerance. Psychological tests yield severe autoaggressive tendencies. The spectrum of skin lesions comprises scratches, ulcers, thermal and caustic burns (e.g. from cigarettes or acids) (**Fig. 7**), hematomas, and many others. Self-manipulation should always be considered when symptoms persist for a very long time in spite of adequate therapy.

Figure 7 Multiple self-inflicted cigarette burns of equal age (1 day) on the left forearm of a 72-year-old psychiatric patient with paranoid ideas (main diagnosis: chronic alcoholism and cerebrovascular encephalopathy).

Munchausen's syndrome is defined as a chronic factitious disorder with systemic malingering observed in adult patients who present themselves to physicians and hospitals with dramatic, but false stories of illness. In assuming the role of physically ill patients, they submit to unnecessary medical procedures including invasive, painful and even dangerous treatment. The factitious diseases are made credible by complaining of symptoms such as abdominal pain, bleeding from body orifices, dermatological changes, and acute neurological manifestations. In contrast, Munchausen syndrome by proxy is a rare form of child abuse in which a parent or other carer, usually the mother, brings a child for medical assessment and care. The proxy either invents a false illness story or really induces an illness (e.g. by nonaccidental poisoning, smothering, or inflicting physical injuries on the child).

Some mentally retarded or disturbed children show special forms of self-destructive behavior such as lip biting, head bumping, and grinding of the teeth. Psychotic patients may be capable of mutilating themselves in a bizarre and painful manner so that the appearance of the injuries is often inconsistent with the outlined criteria of self-infliction. Thus, medicolegal literature reports on self-amputations (fingers, toes, scrotum, penis), but also on enucleation of an eye and severance of the tongue.

Attempted Suicide

An attempted suicide is an action performed with the stated intent of jeopardizing the person's own life. An individual who in desperation gives the appearance of wishing to commit suicide, but lacks the real intention to do so, makes a 'cry for help' or a 'suicidal gesture'. The methods of suicide attempts include the whole spectrum of intoxications and physical trauma: overdosed application of drugs, cutting and stabbing, jumping from heights or in front of a moving vehicle, gunshot injuries, electrocution, nonfatal asphyxia, burning, and jumping into water.

In the context of this article, only suicidal cutting and stabbing is discussed. Suicidal cuts are mostly inflicted by means of knives, razor blades, and other sharp-edged instruments. The flexor surfaces of the distal forearm, the cubital region, and the front of the neck are counted among the preferred 'sites of election'. The presence of linear scars may suggest previous attempts. Suicidal incisions are typically multiple, parallel, superficial, and arranged in groups. The shallow cuts in particular reflect the hesitant or tentative nature of the injuries (**Fig. 8**).

Figure 8 'Suicidal gesture' of a young girl with several parallel, equally shallow, rather scratch-like cuts in the left cubital fossa.

In many suicide attempts the individual abandons the method of cutting the wrists and/or throat after a few trial incisions and turns to another kind of self-destruction which is expected to be more effective. In suicidal cuts of the wrists, the radial and ulnar arteries are severed only in rare cases. The left wrist is the most common target in right-handed persons, but about one half of suicides cut both the left and the right wrist. In suicidal throat cuts, the incisions either pass obliquely across the front of the neck (for instance starting high on the left and ending at a lower level on the right) or the direction is rather a horizontal one. Again, numerous tentative cuts are regarded as a clue to self-infliction.

Suicidal stabbing is less frequent than cutting and usually confined to the precordial region and/or the neck. In both locations the 'classical' pattern is characterized by grouped stabs without concomitant damage of the covering clothes. The wound slits in the skin are often parallel with the pointed end (produced by the cutting edge of the blade) always being on the same side. Similar to suicidal cuts, the existence of multiple trial stabs is accepted as a hallmark of self-infliction.

See also: **Causes of Death**: Sharp Injury. **Clinical Forensic Medicine**: Overview; Defense Wounds; Sexual Assault and Semen Persistence. **Psychology and Psychiatry**: Psychiatry.

Further Reading

Aiken MM, Burgess AW and Hazelwood RR (1995) False rape allegations. In: Hazelwood RR and Burgess AW (eds) *Rape Investigation*, pp. 219–240. Boca Raton: CRC Press.

Bonte W (1983) Self-mutilation and private accident insurance. *Journal of Forensic Science* **28**:70–82.

Crane J (1996) Injury. In: McLay WDS (ed.) *Clinical Forensic Medicine* 2nd edn, pp. 143–162. London: Greenwich Medical Media.

Dotzauer G and Iffland R (1976) Self-mutilations in private accident-insurance cases. *Zeitschrift für Rechtsmedizin* **77**:237–288. (in German)

Dürwald W (1976) Simulation und Selbstbeschädigung. In: Prokop O and Göhler W (eds) *Forensische Medizin*, 3rd edn, pp. 309–330. Stuttgart: Fischer (in German).

Eckert WG (1977) Self-inflicted wounds. In: Tedeschi CG, Eckert WG and Tedeschi LG (eds) *Forensic Medicine*, vol. 1, pp. 496–499. Philadelphia: Saunders.

Eckert WG (1977) The pathology of self-mutilation and destructive acts: a forensic study and review. *Journal of Forensic Science* **22**:242–250.

Faller-Marquardt M, Ropohl D and Pollak S (1995) Excoriations and contusions of the skin as artefacts in fictitious sexual offences. *Journal of Clinical Forensic Medicine* **2**:129–135.

Greenspan GS and Samuel SE (1989) Self-cutting after rape. *American Journal of Psychiatry* **14**:789–790.

Hänsli N (1996) *Automutilation*. Bern: Huber. (in German)

Karger B, DuChesne A, Ortmann C and Brinkmann B (1997) Unusual self-inflicted injuries simulating a criminal offence. *International Journal of Legal Medicine* **110**:267–272.

Knight B (1996) *Forensic Pathology*, 2nd edn. pp. 212–221. London: Arnold.

König HG (1996) Discrimination of sharp injuries, produced in order to feign assaults and such of real cases. In: Saternus KS and Kernbach-Wighton G (eds) *Research in Legal Medicine*, vol. 14, pp. 23–49. Lübeck: Schmidt-Römhild. (in German)

Murray JB (1997) Munchausen syndrome/Munchausen syndrome by proxy. *Journal of Psychology* **131**:343–352.

Püschel K, Hildebrand E, Hitzer K and Harms D (1998) Zur Beurteilung verstümmelnder Hand- und Fingerverletzungen bei Ärzten im Zusammenhang mit privaten Unfallversicherungen. *Versicherungsmedizin* **50**:232–240. (in German)

Rauchfleisch U, Schuppli R and Haenel T (1983) Personality characteristics of patients with dermatological artefacts. *Zeitschrift für Psychosomatische Medizin* **29**:76–84. (in German)

Schulz F, Naeve W, Hildebrand E and Püschel K (1981) Self-inflicted skin injuries and false information to the police. *Aktuelle Dermatologie* **7**:10–16. (in German)

Sexual Assault and Semen Persistence

W Green, University of California Davis, Medical Center, Sacramento, CA, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0441

Introduction

Sexual assault is usually a hidden crime where the only witnesses are the victim and the assailant. For a variety of reasons, even the victim may be unable to provide a detailed account of the assault or the identity of the perpetrator. Often the case history deteriorates into one person's word against another. With such limited initial information, the physical and biological evidence collected from the victim, from the crime scene, and from the suspect will play a pivotal role in the objective and scientific reconstruction of the events in question. Rape is an excellent example of Locard's Principle which states: Every contact leaves a trace. In addition to the hairs, fibers and debris that may be transferred in other types of violent crime, the rapist often leaves behind a personal biological signature that may include blood, saliva, and most importantly, semen. This semen evidence is frequently a cornerstone in the investigation and prosecution of the case.

In any sexual assault investigation, there are four main questions that semen evidence may help answer:

1. Did sexual contact occur? Positive recovery of any of the component parts of semen from the victim is generally considered conclusive proof that sexual contact took place. Recovery of semen from the vagina, however, does not necessarily prove intercourse since other modes of deposition may be responsible.
2. When (or in what time frame) did the sexual contact occur? The interval between semen deposition and evidence collection may be estimated by comparing the specific findings in the case with published norms and maximum reported recovery intervals.
3. Can a specific suspect be included or excluded as the potential source of the semen? Genetic profiles from the evidence material, the victim and the suspect can be developed using conventional serology (antigens and enzymes) and DNA. These profiles can then be compared and a confidence level established to include or exclude a specific suspect.
4. Was the sexual contact consensual or noncon-

sensual? If the victim is beyond the age of consent, semen findings are not helpful. If the victim is underage, then consent is moot and the recovery of semen is consistent with the commission of the crime.

Detection of Semen

The obvious first step in finding semen evidence is interviewing the victim. Initial evidence collection at the crime scene by the law enforcement officer or crime scene investigator should be guided by the victim's description of the event. Items collected for forensic analysis may include clothing the victim wore during or after the assault, bedding, samples from furniture, vehicles, carpeting, or any item the victim may have used to clean up with after the assault (tissue, washcloth, etc.).

Additional victim-guided sampling will be obtained in the course of the forensic medical examination. If not already retained, the 'assault' clothing can be collected. Highest priority is directed to items the victim believes contain semen, areas of visible staining, and any clothing item that was nearest the genital area. The examiner will also obtain samples from the victim's body. Again, any area the victim believes was in contact with semen will be carefully swabbed. Any material in the victim's hair (pubic, head or body) will be clipped out and retained. Scanning the victim with an ultraviolet light source may reveal the fluorescence of dried semen, which was not visible in plain light. The examiner must collect samples from any body cavity or opening where sexual contact was attempted or completed. Unfortunately, the victims are rather inaccurate in their perceptions about semen deposition in or near body cavities. Possible explanations include the severe psychologic stress of the assault and the potential release of small, but detectable, amounts of semen in the pre-ejaculate emission. Semen has also been recovered in cases where the victim believed the assailant wore a condom, so samples should be obtained despite the history of condom use.

The process of collecting biological samples from the victim usually involves the application of moistened swabs to areas of possible dried stains and the insertion of dry swabs (under direct visualization) into body cavities to absorb any liquid secretions. Some jurisdictions prefer expectorated liquid saliva samples for oral cavity sampling or vaginal aspiration (or lavage) after vaginal penetration.

The proper handling and packaging of potential semen evidence cannot be overstressed. Any 'moist' or 'wet' evidence must be thoroughly air dried. In the case of swabs, this requires placement in a stream of cool air for at least 60 min. Potential biological evidence must be packaged in containers that allow air circulation; never in plastic bags or sealed, nonporous tubes, jars, or boxes. Chain of custody must be meticulously documented and adhere to the prevailing local policies. Transportation, deposition, and storage of evidence must also strictly follow local protocol.

Evaluation of Potential Semen Evidence

Evidence material presented to the forensic laboratory for analysis of possible semen will usually consist of swabs taken from various locations on or in the victim and objects (clothing, fabric, matted hair, etc.) that may contain semen. Every laboratory will have its own specific protocol and sequence for analyzing the evidence but the general approach is similar (**Table 1**).

Screening tests

Ultraviolet light To help identify occult semen stains, the evidence may be scanned with an ultraviolet light (or alternative light source). Dried semen may fluoresce under UV illumination and highlight stains that were previously invisible thus directing the criminalist to sample areas of higher potential yield.

Presumptive chemical tests for semen Substances within the evidence swabs or stains that allow the

Table 1 Evaluation of potential semen evidence in sexual assault investigation

Screening tests
Ultraviolet light scanning
Presumptive chemical tests for semen
Choline (Florence test)
Acid phosphatase (Brentamine test)
Confirmatory tests
Spermatozoa
Motile
Non-motile
Proteinase K enhancement
Fluorescence in situ hybridization (FISH)
Noncellular semen markers
Acid phosphatase
p 30
Individualization of semen evidence
Blood group typing
ABO(H) system
Lewis antigens
Enzyme typing
Phosphoglucomutase (PGM)
Peptidase A (Pep A)
DNA

identification and/or the individualization of semen must be solubilized and extracted for analysis. Presumptive testing for semen helps to confirm the sampling location as a potentially positive site. The two commonly used presumptive tests for semen are both colorometrics. The Florence test identifies the presence of choline and the Brentamine test detects acid phosphatase. Both of these assays are qualitative.

Confirmatory testing

Confirmatory testing for semen involves solubilization of evidence material followed by appropriate vortexing and centrifugation, which yields a supernatant and a cell pellet. The cell pellet is used primarily to detect spermatozoa and for DNA analysis. The supernatant portion is most helpful to detect noncellular markers in semen when sperm are not detected and to develop genetic profiling or grouping.

Spermatozoa Identification of one or more intact spermatozoa is conclusive proof of the presence of semen and hence affirms sexual contact. The condition and collection site of the sperm may be very helpful in defining the time interval between deposition and collection (also known as the postcoital interval or PCI).

Motile sperm The forensic medical examiner has the unique opportunity to observe motile sperm collected from the victim during the evidentiary exam. Recovery of motile sperm has only been reported from samples taken from the vagina or cervix. The technique requires the preparation of a 'wet mount' slide (vaginal or cervical swab sample placed on a slide with a drop of saline plus a cover slip) and examined with a phase-contrast microscope. The maximum reported recovery times are shown in **Table 2**.

Nonmotile sperm Although nonmotile but morphologically intact spermatozoa may be seen on the wet mount slide performed during the forensic medical examination, the identification of nonmotile sperm is usually based on the stained smears prepared by the forensic laboratory from the evidence swabs. Commonly employed staining methods include Oppitz (Christmas-tree-stain), hematoxylin and eosin (H & E) and Papanicolaou (pap smear). The slides are usually evaluated by plain light microscopy at 400 ×. Some laboratories score the concentration of spermatozoa seen. This is usually reported as a range from 'few' (less than five sperm per field) to '4+' (many in every field). If the sample has been contaminated by other bodily fluids (saliva, vaginal secretions, etc.), epithelial cells and cellular debris may make the identification of spermatozoa more problematic. Selective degradation of epithelial cells and debris may be accomplished by treating the cell extract with a mixture of proteinase K and sodium dodecyl sulfate before staining and microscopic examination.

Table 2 summarizes the maximum reported recovery times for motile and nonmotile sperm collected from various anatomic locations. These 'record' times were compiled from multiple published studies and represent a variety of methodologies. Recent investigators have been able to compare sperm concentrations (few to 4+) found on the stained smears to the time since offense to refine estimates of probable postcoital interval.

A number of factors may influence the recovery of spermatozoa. Because the number of sperm collected per swab may be very low, most authorities suggest obtaining multiple swabs from the same site to maximize recovery and thus provide more forensic material for analysis. Victim position following the assault plays an important role in the amount of vaginal spermatozoa available for sampling. Gravity drainage creates significant loss of vaginal evidence if the victim is upright after deposition. Conversely, if the victim remains recumbent following the act, recovery is likely to be greater. This has been a legitimate criticism against comparing volunteer studies (more likely recumbent) to casework studies (more often upright). If only a small number of sperm are recovered from the vagina, the possibility of sperm deposition from a contaminated object (e.g. finger) must be considered as an alternative explanation to penile penetration.

The detection of sperm on anal swabs must be interpreted cautiously. Recovery of anal sperm from a male sodomy victim or from a female sodomy victim without any detectable vaginal sperm is consistent with anal sexual contact. Difficulty arises when the female victim has both vaginal and anal sperm. Drainage of vaginal material may contami-

Table 2 Maximum reported recovery times for spermatozoa collected from body cavities in living sexual assault victims

	Vagina	*Cervix*	*Mouth*	*Rectum*	*Anus*
Motile sperm	6–28 h	3–7.5 days	—	—	—
Nonmotile sperm	14 h to 10 days	7.5–19 days	2–31 h	4–113 h	2–44 h

nate the anal area. Sampling technique, specific labeling of 'anal' and 'rectal' samples and sperm concentrations all bear on the interpretation of the individual case.

The recovery of spermatozoa from, in or around the mouth is generally low. Some authorities now advocate oral sampling by expectorated liquid saliva to improve the yield over the standard swab technique.

Fluorescence in situ hybridization A new molecular cytogenic analysis method offers an innovative approach to improve the identification of male cells from sexual assault evidence materials. Fluorescence *in situ* hybridization (FISH) uses a Y chromosome specific DNA probe to identify Y-bearing (male) cells. This technique sensitively identifies not only spermatozoa, but also nonsperm cells (epithelial and segmented neutrophils) of male origin. Although forensic experience with this technique in sexual assault cases is limited, it offers a sensitive and specific alternative to conventional methods for the primary or confirmatory detection of spermatozoa. It also offers the additional potential for confirming male–female contact when sperm and other semen markers are absent.

Noncellular semen markers When evaluation of the potential semen evidence yields a positive presumptive test or a negative presumptive test but high clinical suspicion and the analysis fails to identify spermatozoa, further testing is required. The objective is to detect the presence of markers that are specific and unique to seminal plasma but independent of sperm cell presence. The two most commonly employed semen constituents are acid phosphatase and the glycoprotein, p30.

Acid phosphatase Acid phosphatase is not a single enzyme but an array of related isoenzymes from a variety of sources. The forensic interest in acid phosphatase in the evaluation of sexual assault evidence is based on the fact that acid phosphatase activity in human semen is 500–1000 times greater than in any other normal bodily fluid. Unfortunately, the use of acid phosphatase as a marker for semen is compromised because the vagina is also a source of the same type of acid phosphatase. Since seminal and vaginal acid phosphatases cannot be reliably discriminated qualitatively, the only approach to differentiating azoospermic semen from vaginal secretions is by quantitative analysis. Finding a 'significantly' elevated acid phosphatase level is consistent with the presence of semen.

A number of factors, however, complicate the interpretation of any given sample. Considerable variation in baseline endogenous vaginal acid phosphatase levels occurs not only between individuals, but also within a single individual. Pregnancy, phase of the menstrual cycle, bacterial vaginosis, use of certain feminine hygiene products or other contaminants may produce elevated endogenous acid phosphatase values. Measurable acid phosphatase activity in vaginal samples declines after semen deposition because of drainage and dilution from vaginal secretions. The rate of decline is quite valuable. There is little consensus in the literature regarding a reliable threshold level to consistently separate endogenous vaginal acid phophatase from seminal. There is also no agreement on the time frame in which acid phosphatase analysis is useful for forensic purposes following deposition. Reports vary from as short as 3 h to as long as 72 h. Because physiologic, temporal, quantitative and methodologic variables all potentially affect measured values, the interpretation of a specific specimen should be in the context of the database and experience of the forensic laboratory performing the test.

p30 The glycoprotein, p30 (also known as prostate specific antigen or PSA), is derived from prostate epithelial cells and is found in seminal plasma, male urine, and blood. p30 has not been found in any female body tissue or fluid. There are no significant differences in p30 recovery from vasectomized men compared to nonvasectomized controls. The finding of any p30 in evidence material confirms the presence of semen. A positive p30 analysis reliably identifies semen regardless of whether acid phosphatase is elevated or spermatozoa are detected.

As with other markers, the level of p30 declines after deposition in the vagina. The disappearance is log-linear and fairly predictable. The mean time to reach the lower limits of vaginal detection is 27 h with a range of 13–47 h. p30 thus provides an additional point of reference with which to estimate the postcoital interval. p30 is quite durable in dried stains and has been detected up to 10 years later in material stored at room temperature.

Monoclonal antibody mouse antihuman semen-5 (MHS-5) Another semen-specific antigenic protein has been detected but is not widely used in the forensic evaluation of rape. MHS-5 is produced in the seminal vesicle epithelium and is not found in any other bodily fluid besides semen and has no cross-reactivity with other body fluids. The mean vaginal decay time has not been established. Although forensic experience with the technique is limited, it may potentially be a viable adjunct in the arsenal of confirmatory semen tests.

Individualization of semen stains

Once evidence material has been shown to contain semen, the next step is to individualize the semen by developing a genetic profile of the donor. This genetic profile can then be compared with the genetic profiles of the victim and the suspect. The suspect can thus be included as the possible assailant or excluded from consideration. The more definitive the genetic typing, the greater the probability of excluding innocent suspects and the more restricted the guilty suspect group.

Semen contains many polymorphic protein markers but only a few are amenable to forensic analysis because most are only recovered in amounts too small for reliable detection or evaluation. Conventional forensic serology is usually limited to soluble blood group antigens (ABO (H) system and Lewis antigens), phosphoglucomutase (PGM) and peptidase A (Pep A).

It would not be an overstatement to say that the recent explosion in DNA technology is revolutionizing forensic science. Given the fact that DNA is both highly polymorphic and extremely stable, the forensic laboratory now has the capability to employ individualizing techniques on minute evidence samples with extraordinarily high sensitivity and discriminatory potential.

Blood group typing The human red cell membrane contains hundreds of genetically determined antigens. The best studied and most important forensically is the ABO group. Closely associated with the ABO system is the 'H' antigen, which is a precursor of the A and B antigens. Although the H and ABO loci are not genetically linked, their functional relationship has prompted the convention of treating these antigens as a single group.

Although most red cell antigens are bound to the cell membrane, the ABO(H) antigens are soluble and can be secreted into bodily fluids (including saliva, semen and vaginal secretions). This ability to secrete soluble antigens is under the control of a pair of genes, Se and se. With Se being dominant, homozygous (Se Se) and heterozygous (Se se) individuals are 'secretors' and homozygous (se se) individuals are 'nonsecretors". About 80% of the population are secretors. Clinically, the secretor status of an individual can be checked by comparing that individual's ABO blood type with the presence or absence of the same ABO(H) antigens analyzed from a saliva sample.

The detection of Lewis antigens in the blood is another method of establishing secretor status. Lewis (a− b+) phenotypes are secretors and Lewis (a+ b−) are not. The secretor status of Lewis (a− b−) individuals are variable. In addition to secretor information, Lewis antigen phenotypes have different distributions in various racial groups.

The application of blood group typing in sexual assault cases requires the comparison of blood group substances recovered in the evidence material with those of the victim and the suspect (**Table 3**). The secretor status of both is also important for the proper interpretation of results. A rough quantitative assessment of the amount of antigen deposited by the assailant may be inferred by the concentration of spermatozoa found on the vaginal smear (few to 4+). The estimation of foreign antigen amounts may be used with the measured amounts of recovered antigens to further refine the interpretation of blood group results. As with the other marker assays, temporal, quantitative, methodological and physiologic variables may limit the usefulness of the blood group contribution to the genetic profile.

Despite these limitations and the substantial discriminatory power of DNA, conventional blood group typing retains a significant role in the evaluation of sexual assault evidence. Traditional grouping is cheap, fast and universally available. ABO blood grouping may be superior to DNA analysis for typing saliva or semen that contains few or no sperm. Seminal blood groups have been detected in the vagina up to 21 h after deposition. Seminal blood groups are rarely recovered from the mouth or anus/rectum.

Enzyme typing Phosphoglucomutase (PGM) and peptidase A (Pep A) are the two enzyme markers commonly used in the genetic profiling of semen evidence in sexual assault cases. These enzymes are found in semen and vaginal secretions regardless of ABO type or secretor status. Phosphoglucomutase is polymorphic in all populations and can be subdivided into ten distinct subgroups. Although peptidase A is polymorphic in many racial groups, it is most commonly used as a discriminator in cases in which the perpetrator is thought to be Black.

The measurable activity of both PGM and Pep A decline rapidly in the postcoital vagina with PGM usually dropping below threshold by 6 h and Pep A not usually recovered after 3 h. An additional

Table 3 ABO blood types and antigens in the forensic evaluation of sexual assault.

Victim's ABO phenotype	*Expected antigens from the victim*	*Foreign antigens from the assailant*
O	H	A and/or B
A	A and H	B
B	B and H	A
AB	A and B and H	None

problem with PGM is that semen contaminated with saliva can exhibit aberrant PGM isoenzyme patterns.

DNA profiling The primary advantage of DNA profiling in sexual assault investigations is its ability to accurately individualize semen that contains only minimal numbers of spermatozoa. DNA can also be used to differentiate multiple donors in mixed stains whether there are mixed sources of the same bodily fluid or different bodily fluids contributed by different individuals. DNA technology also enhances other techniques like polymerase chain reaction, restriction fragment length polymorphism, ABO genotyping and fluorescence *in situ* hybridization of Y-bearing male cells. The already long list of DNA advancements is growing rapidly. As the use of DNA techniques becomes cheaper, more widely available and more uniformly accepted by the courts, the greater will be the role of the forensic laboratory in the investigation and prosecution of sexual assault cases.

Postmortem Detection of Semen

Very little has been published about the postmortem recovery of semen from rape–homicide victims. One report found morphologically intact spermatozoa in the vagina of a rape–homicide victim who had been dead 16 days. A more recent report describes the recovery of intact spermatozoa from the vagina of a homicide victim at 34 days postmortem. This victim also had detectable p30 on the vaginal swab. No information was available in either report as to the time interval between deposition of semen and death.

A few anecdotal reports have been published describing the recovery of sperm from dead bodies. In these cases there was no history of sexual assault, but swabs were obtained at autopsy as a 'precaution'. Little or no information was available about the time of deposition relative to the time of death. The longest reported postmortem vaginal recovery was 3–4 months. One body with vaginal sperm was immersed in water up to 2.5 weeks. One mummified corpse yielded vaginal sperm 5–6 weeks postmortem. A mouth swab was positive 4 days after death.

Failure to Recover Semen

When the history anticipates the recovery of semen from a rape victim but the forensic evaluation fails to find any semen evidence, there are a number of potential explanations besides the obvious suggestion of a false complaint (**Table 4**). No semen will be recovered if none was deposited (sexual dysfunction, condom use). Mechanical elimination (drainage, hygiene activities), biological elimination (degradation) or physiologic dilution (or any combination) may yield negative results.

Table 4 Factors that may limit the detection or recovery of semen from the sexual assault victim

No seminal constituents recovered
Sexual dysfunction in the assailant
Condom use
Time delay between the assault and the exam:
Drainage of vaginal contents
Degradation of seminal secretions
Dilution by endogenous components
Victim hygiene (douching, bathing, gargling, etc.)
Physiologic activity (urinating, defecating, menses)
No spermatozoa recovered
Impaired spermatogenesis:
Germinal aplasia/hypoplasia
Hormonal deficiencies (pituitary, thyroid, gonadal)
Testicular irradiation
Drugs/toxins
Neoplasm
Varicocele
Infection (mumps, tuberculosis, gonorrhea)
Elevated scrotal temperature
Chronic alcoholism
Depleted stores (frequent ejaculation may create transient azoospermia)
Impaired delivery
Vasectomy
Trauma
Congenital abnormalities

Seminal plasma markers may be present, but many factors may preclude the recovery of spermatozoa. Impaired spermatogenesis, depleted stores, or impaired delivery may all play a role in the production of azoospermic semen.

See also: **Serology**: Overview; Blood Identification. **Deoxyribonucleic Acid**: Restriction Fragment Length Polymorphism.

Further Reading

Allard JE (1997) The collection of data from findings in cases of sexual assault and the significance of spermatozoa on vaginal, anal, and oral swabs. *Science and Justice*, 37:99–108.

Baechtel FS (1988) The identification and individualization of semen stains. *Forensic Science Handbook* **2**:347–349.

Collins KA, Rao PN and Hayworth R *et al* (1994) Identification of sperm and non-sperm male cells in cervicovaginal smears using fluorescence in situ hybridization: applications in alleged sexual assault cases. *Journal of Forensic Science* **39**:1347–1355.

Dahlke MB, Cooke C, Cunnae M et al (1977) Identification of semen in 500 patients seen because of rape. *American Journal of Clinical Pathology* **68**:740–746.

Davies A (1979) Discussion of 'Spermatozoa in the anal canal and rectum and in the oral cavity of female rape victims.' *Journal of Forensic Science* **24**:541–542.

Davies A, Wilson EM and Silverman AG (1974) The persistence of seminal constituents in the human vagina. *Journal of Forensic Science* **3**:45–55.

Enos WF and Beyer JD (1979) Spermatozoa in the anal canal and rectum and in the oral cavity of female rape victims. *Journal of Forensic Science* **23**:231–233.

Keating SM and Allard JE (1994) What's in a name? Medical samples and scientific evidence in sexual assaults. *Medicine Science and the Law* **43**:187–201.

Morrison AI (1972) Persistence of spermatozoa in the vagina and cervix. *British Journal of Veneral Disease* **48**:141–143.

Perloff WH (1964) Steinberger: in vivo survival of spermatozoa in cervical mucus. *American Journal of Obstetrics and Gynecology* **88**:439–442.

Pollak OJ (1943) Semen and seminal fluids. *Archives of Pathology* **35**:140–187.

Rao PN, Collins KA, Geisinger KR *et al* (1995) Identification of male epithelial cells in routine postcoital cervicovaginal smears using fluorescence in situ hybridization. *American Journal of Clinical Pathology* **10**:32–35.

Rupp JC (1969) Sperm survival and prostatic acid phosphatase activity in victims of sexual assault. *Journal of Forensic Science* **14**:177–183.

Sharpe N (1963) The significance of spermatozoa in victims of sexual offenses. *Canadian Medical Association* **89**:513–514.

Silverman EF (1977) Persistence of spermatozoa in the lower genital tracts of women. *American Journal of Clinical Pathology* **68**:107.

Silverman EM and Silverman AG (1978) The persistence of spermatozoa in the lower genital tracts of women. *Journal of the American Medical Association* **240**:1875–1877.

Wallace-Haagens MJ, Duffy BJ and Holtrop HR (1975) Recovery of spermatozoa from human vaginal washings. *Fertility and Sterility* **26**:175–179.

Willot GM and Allard JE (1982) Spermatozoa – their persistence after sexual intercourse. *Forensic Science International* **19**:135–154.

COMPUTER CRIME

H Henseler, Netherlands Organization of Applied Scientific Research, Institute of Applied Physics, Delft, The Netherlands

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0444

Introduction

The days of Sherlock Holmes are over. Many things have changed in the world around us. An ever-increasing flow of information is sent through computers, fax or mobile phone communications. Computer systems and electronic organizers are replacing paper administrations and diaries and criminals operating in worldwide organizations are using this kind of modern equipment. Figures show that the number of computers and digital storage media that are seized by police is continually increasing. Intercepting data communication in wiretaps used to be restricted to fax messages. Today, a wiretap on a high-speed modem connected to an internet provider will give a large variety of message formats. Analyzing this kind of evidence requires a fundamental change of methods, which has resulted in the rapid development of forensic computer science. This article addresses basic types of computer crime and subsequently discusses the application of forensic computer science which is applicable in a broad range of crime sorts, i.e. organized crime, drug trafficking, violent crimes, fraud and also computer crime.

Computer Crime

There is no exact global definition for computer crime, sometimes also referred to as computer-related crime. Rather, functional definitions have been used to describe the phenomenon. It can involve many criminal activities that are traditional in nature, such as theft, fraud, forgery and mischief, all of which are generally subject to criminal sanctions. The computer has also created a host of potentially new misuses or abuses that may, or should, also be criminal.

Five common types of computer crime have been identified and are summarized here.

Fraud by computer manipulation Intangible assets represented in data format, such as money on deposit

or hours of work, are the most common targets of computer-related fraud. Modern business is quickly replacing cash with deposits transacted on computer systems, creating an enormous potential for computer abuse. Credit card information, along with personal and financial information on credit-card clients, have been frequently targeted by the organized criminal community. The sale of this information to counterfeiters of credit cards and travel documents has proven to be extremely lucrative. Assets represented in data format often have a considerably higher value than traditionally targeted economic assets, resulting in potentially greater economic loss. In addition, improved remote access to databases allows the criminal the opportunity to commit various types of fraud without ever physically entering the premises of the victim.

Computer forgery Where data are altered in respect of documents stored in computerized form, the crime is forgery. In this and the above examples, computer systems are the target of criminal activity. Computers, however, can also be used as instruments with which to commit forgery. They have created a new library of tools with which to forge the documents used in commerce. A new generation of fraudulent alteration or counterfeiting emerged when computerized color laser copiers became available. These copiers are capable of high-resolution copying, the modification of documents and even the creation of false documents without the benefit of an original, and they produce documents whose quality is indistinguishable from that of authentic documents except by an expert.

Damage to or modification of computer data or programs This category of criminal activity involves either direct or covert unauthorized access to a computer system by the introduction of new programs known as viruses, 'worms' or logic bombs. The unauthorized modification, suppression or erasure of computer data or functions with the internet to hinder normal functioning of the system is clearly criminal activity and is commonly referred to as computer sabotage. Computer sabotage can be the vehicle for gaining economic advantage over a competitor, for promoting the illegal activities of ideologically motivated terrorists or for stealing data or programs (also referred to as 'bitnapping') for extortion purposes. In one reported incident at London, Ontario, in 1987, a former employee of a company sought unsuccessfully to sabotage the computer system of the company by inserting a program into the system that would have wiped it out completely.

Unauthorized access to computer systems and service The desire to gain unauthorized access to computer systems can be prompted by several motives, from simple curiosity, as exemplified by many hackers, to computer sabotage or espionage. Intentional and unjustified access by a person not authorized by the owners or operators of a system may often constitute criminal behavior. Unauthorized access creates the opportunity to cause additional unintended damage to data, system crashes or impediments to legitimate system users by negligence.

Unauthorized reproduction of legally protected computer programs The unauthorized reproduction of computer programs can mean a substantial economic loss to the legitimate owners. Several jurisdictions have dictated that this type of activity should be the subject of criminal sanction. The problem has reached transnational dimensions with the trafficking of these unauthorized reproductions over modern telecommunication networks.

In modern information society computer crime should be identified as a serious threat and as such it is an emerging challenge for law enforcement. Due to its nature, special technical countermeasures are required that are not restricted to the forensic analysis of computer evidence but also require advanced law enforcement techniques for digital investigation. For example, the internet requires new techniques and procedures to enable the police to enforce laws on the electronic highway. Also public prosecutors should be aware of new telecommunication and computer technology. Typical examples of computer crime are hacking (i.e. breaking into a computer system), phreaking (i.e. manipulating the phone system), software piracy and spreading malicious code (i.e. computer viruses). These emerging crimes require a specialized computer crime law such as the one that was introduced in The Netherlands in 1993. Among other things, this particular law makes unauthorized access to any password-protected computer system illegal. Other examples such as child pornography on the internet, money laundering, illegal gambling on the internet, credit card fraud etc., should not be viewed as computer crime but as digital equivalents of their 'real world' counterparts. Traditional laws still apply to these crimes but new techniques are required to investigate these matters.

Forensic Computer Science

Although computer crime is viewed as a serious threat it is not the driving force behind the development of forensic computer science. The investigation

of traditional crimes like murder, fraud and drug trafficking require new techniques since organized crime is using modern communication techniques (e.g. e-mail, fax, GSM) and computerized administrations (e.g. personal computers (PCs), electronic organizers). We must conclude that the introduction of computers in many areas of our society has led us to a point where traditional methods for investigating and analyzing evidence are no longer adequate in the new information society.

Here forensic analysis of computer evidence (or in short computer forensics) is subdivided in three major categories:

1. embedded computer systems;
2. open computer systems and networks;
3. communication systems.

This subdivision largely reflects the nature of the items that are seized or intercepted and submitted for further forensic investigation. Typically, consumer electronics belong to the first category, for example an electronic organizer or a mobile phone. Accessing information in embedded computer systems requires analysis of the hardware using special equipment. The open systems refer to items such as computers and storage media that adhere to open standards and that are accessible via software rather than hardware, for example MS-DOS, Wintel computers or plain standard storage media such as IDE hard drives, computer tapes and disks. Not only data but also software in open systems, may contain interesting traces of evidence and should be investigated. Traces of evidence are found in both embedded systems and open systems by investigating physical items, e.g. electronic organizers, PCs, disks, tapes etc. In contrast, items found in communication systems are not physical but are found in, for example, digital transmissions.

Embedded computer systems

Embedded systems are found in the field of consumer electronics. This is a field that has experienced tremendous new development over the past three years, especially the explosive growth of the digital telecommunications, e.g. mobile GSM (DCS800 in Europe or DCS1800 in the US), the large-scale introduction of smart card applications and the revival of the electronic organizer, all of which have greatly influenced the development of new forensic techniques. Electronic equipment has become more user friendly and able to store information which may be interesting when searching for traces of evidence. Hardware investigations are further complicated by the continuing miniaturization in combination with increased functionality, e.g. organizer, phone, fax and e-mail.

Investigating evidence traces in electronics requires a hardware approach. Skilled technicians with a broad knowledge of computer architectures are essential to analyze pieces of hardware when looking for memory locations. Sometimes, even removing the back panel of an electronics device and putting it back together requires special equipment and experience. This is even truer for encapsulated integrated circuits (ICs), for example in the case of a smart card. The introduction of multifunctional smart card applications (i.e. for payment, insurance, and phone) simply begs for new forensic methods. And this will become more interesting once biometrics (e.g. fingerprint, hand) are used so that digital traces have larger potential to deliver important evidence. Hardware investigation of smart cards and ICs requires either chemical solvents or strong acids that leave the electronics undamaged but remove surrounding epoxy in a fast and clean way. Moreover, investigation of microelectronics requires miniature equipment, and a microscope is needed to analyze the inside of an IC. Measured signals can subsequently be analyzed using standard methods or with special-purpose software. Manual analysis is not feasible owing to the high clock speeds which yield millions of samples per second. However, analysis is still essential. Sometimes it is sufficient to examine data only and look for particular information. In other cases it may be necessary to analyse the firmware (software in the embedded hardware) before a good interpretation or recovery of the information is possible. Since embedded hardware mostly operates with special-purpose processors using an unknown machine instruction set, forensic analysis becomes a difficult and laborious task.

One might ask what is the use of analysing an embedded system in such hardware detail? It is tempting to think that a good manual will yield all the information that is stored in the equipment. For example, both mobile phones and electronic organizers can be accessed using the keypad. However, things are not what they seem to be. First, making a memory dump is more reliable and faster than manually copying all information. Secondly, memory may contain traces of information that has been erased by the user and which is not normally accessible. Removing information from an index does not always mean that the information has been erased in the memory. In case of broken equipment, making a memory dump may be a final resort to recover any information at all. Finally, in many cases, the computer will keep more information in records that the user can access. For instance, it may be possible to discover in which chronological order information was entered into the system.

Open computer systems and networks

The major changes in the forensic investigation of open systems have been caused by the fast development of the PC. A well-known PC type is the IBM compatible PC with the MS-DOS operating system. In the 1980s, Unix was the main operating system on minicomputers and workstations in computer networks. PCs were used mainly as stand-alone systems. In the late 1980s and early 1990s, the Novell network enabled PCs to be connected to a server in a network. Three years ago the MS-DOS operating system played a central role in the forensic investigation of storage media and operating systems. With the arrival of first Windows for Workgroups 3.11 and later Windows 95 and Windows NT 3.51 a new trend was set that greatly influenced the forensic investigation of open systems. Operating systems are without exception graphical, new file systems are being used (FAT-32 and NTFS) and operating systems are object oriented. A thorough analysis of a PC requires new methods, new software and a new approach. The shift from stand-alone to network PCs further complicates the analysis of stored information.

In contrast to digital electronics in embedded systems, PCs and storage media in open systems are not opened with screwdrivers or probed with logical analyzers. Instead they are probed and processed with software: software to copy and search information and software to write special programs that access information that would otherwise remain inaccessible. Most investigations have to deal with MS-DOS and increasingly with Windows computers as well as removable storage media, e.g. computer tapes and disks that may contain gigabytes of information. Knowing how to approach the open system is essential. Removable media can not only have different physical formats (i.e. a 5.25 inch versus a 3.5-inch disk) but they can also have different logical formats. Beware of the 3.5-inch disk that turns out to be used on a machine writer rather then a PC. Computer centers at large companies and universities have experience of dealing with a large variety of different digital media. The equipment that has been developed for them is expensive but can speed-up the forensic investigation considerably. Accessing zeros and ones is the first step. The next step is accessing the actual information that is represented by the zeros and ones. Envision a relational database with 100 data tables. Accessing the information in one table may not be sufficient. Accessing the entire relational database could be essential to find the required evidence. In cases where relations are hard-coded in software this will probably imply that not only the data but also the original software and perhaps the operating system must be restored.

So far, only the analysis of storage media has been discussed. The name 'open system' may give the impression that open systems are not or cannot be secured. In reality, open systems support a large variety of security mechanisms because they are open. Open Standards, for example the Wintel standard, enable many different hardware and software manufacturers to incorporate their own security module. PC security software restricts file access and encrypts information before it is stored. In order to analyse traces of information it is necessary to analyse these security modules and to study their effect on the information. In some cases the operation of programs has to be analyzed in detail, disassembling every single instruction, in order to bypass the activated security mechanisms. This is better known as reverse engineering. Reverse engineering can also be quite useful in case a fast implementation of an encryption program is needed to launch a brute-force attack on a security mechanism. Speed is not always essential. In some pin-code systems only ten thousand trials are necessary to try every possible key combination. With a machine that does two tries per minute, it will take one week on average to break the code. Obviously, such an attack will not succeed for security systems that restrict the number of trials.

Both analysis of information storage and software require knowledge of the underlying operating system. Analysis of complex operating systems, e.g. Unix, Windows NT and even Windows 95, is essential to computer forensics. The term 'Unix Forensics' has been introduced to emphasize the complexity of the forensics that is required to capture data in computers running Unix. These systems will register actions of identified users in various ways. Advanced Windows programmers know that information concerning the use of the system is stored in the system registry and countless configuration files on the hard drive. Moreover, systems running multitasking operating systems function as nodes in a network may contain information residue of communications from the network. Forensic investigation of computer networks is essential to isolate traces of computer hackers that have penetrated the network security and have entered a computer system. With the current explosion of internet connectivity and the increasing capabilities for electronic payment as well as e-commerce, computer hacking is once again a growing item that requires thorough investigation. A detailed discussion of hacking is beyond the scope of this article and the interested reader is referred to the further reading list.

Communication systems

In a relatively short time the fax has become a very popular means of communication. It has been and perhaps still is the most important form of data communication encountered in criminal investigations. In the early days, fax messages were a subject for forensic computer science research. However, the highly standardized communication protocols for fax transmissions soon led several commercial companies to develop turnkey solutions for intercepting fax messages. These solutions have been integrated in modern wire-tapping equipment. Fax communication was no longer a priority in forensic computer science and fax interceptions that were recorded without proper equipment were considered to be an acceptable loss. With the introduction of high-speed modems the need for analyzing data communication has once again gained top priority. Until a few years ago, forensic analysis of modem communication was restricted to rare cases of computer hacking or software piracy in which suspects used modems to connect to Bulletin Board Systems (BBS) and dial-up computers. As the internet and world wide web (WWW) have experienced an exponential growth and because high-speed computer modems have enabled serious on-line applications, modem communication plays a central role in the communication of information and it influences the development of forensic computer science.

In contrast with fax communication, internet communication follows 'open standards' that change every week. The internet can deliver multimedia content and potentially provides much more bandwidth than fax to communicate information in an organization. Internet and WWW protocols have become more complex and may require an effort that sometimes even out ranges reverse engineering of software. The OSI-layer model varies from physical transport layer to application layer. Every layer adds its own information by assembling new packets of information that have to be routed over the physical infrastructure. Well-known protocol keywords, e.g. TCP/IP, ATM, PPP, POP, HTTP, HTML, SSL etc., give a new dimension to forensic computer science. The analysis of layer after layer can be a laborious task even when the used protocols are known. Sometimes, analysis may yield interesting information about the sender and receiver. Also, information may be lost. In case compression techniques were used, a detailed analysis of the partial data is required when attempting to recover as much information as possible. This requires advanced knowledge of network protocols and information-encoding techniques in combination with good programming skills to convert raw data into readable (or visible or audible) information.

Currently, new techniques age quickly. In the upper layers of the OSI model (application oriented) new protocols or message formats are being introduced at high speed. This has become possible through the introduction of plug-in technology that allows WWW browsers to plug-in new software to deal with new data formats. Developments such as Java, COM (common object model) and the distributed version (DCOM) allow an almost instant spreading and online installation of Java applets and Active-X software components in computer systems. Today, the already enormous flow of information is increased even further by so-called push technology that delivers information automatically to the user. The stand-alone application of today is the network application of tomorrow. Perhaps new standards as DCOM will have the greatest impact on forensic analysis when data objects are virtually distributed over an entire computer network still enabling a single point of access. Being able to preserve all information stored in a distributed compound document storage will make analysing clusters in a fragmented FAT-16 file system look like child's play.

Knowledge of data communication and information encoding is not sufficient. To be able to intercept messages, knowledge of the telecommunication infrastructure is essential. This is true for both the fixed and the mobile infrastructure. Service providers will compete by offering different kinds of services to their customers. In mobile telephony, both the handset and the infrastructure may contain information that can prove to be essential evidence, for instance, pre-set numbers, voice-mail and dialled numbers.

Cryptography

Over the past few years cryptography has become of increasing importance to forensic computer science. First, the increased speed of personal computers has enabled the use of cryptography in virtually any security application. Secondly, there is more privacy awareness and users are becoming more and more security minded when dealing with computer networks and e-mail. In such applications, cryptography can be a serious obstacle for a forensic investigation. It can be safely assumed that this problem will further emerge once stronger encryption becomes commercially available. From the point of view of prevention this can be considered good. Too often electronic information can be easily read by others and privacy of customers is not always very well protected. More encryption would be an advantage and should perhaps be made obligatory from a privacy point of view. However, cryptography in the hands of a criminal organization is a dangerous weapon that can be a

serious threat to society. Either key-recovery schemes or key-escrow schemes in combination with a trusted third party (TTP) system will hopefully present a solution for this dilemma. The main condition will have to be that only a judge can order the decryption of encrypted information.

Cryptography also plays another part in forensic computer science. Many results of forensic investigations will be attached to the report in digital format, for example on a computer disk. To guarantee the authenticity of the digital results, a cryptographic hash function is used. A hash function is a mapping from a binary string of arbitrary length to a binary string of some fixed length, called hash value. The hash value is mentioned in the report. This hash value represents a numerical message digest that is uniquely related to the contents of the message by a cryptographic algorithm. For example, the Secure Hash Algorithm (SHA-1) is one of the state-of-the-art algorithms that has been developed by NIST and can be found on the internet as standard FIPS 180-1. The Forensic Computer Science department of The Netherlands Forensic Science Laboratory has implemented the SHA-1 algorithm that has been certified by NIST. This implementation is used to secure digital evidence that is distributed with the report of the forensic investigation. It has been reported that even CD-ROMs can be altered which underlines the need for securing digital evidence by cryptographic means. We can expect to see similar applications of cryptography for authentication in the near future. It is possible that computer forensics will also have to answer questions regarding the strength of such digital signatures.

See also: **Forgery and Fraud**: Payment Cards. **Document Analysis**: Forgery/Counterfeits.

Further Reading

Bates J (1997) The fundamentals of computer forensics. *International Journal of Forensic Computing* **January**.

Carter DL and Katz AJ (1996) Computer crime an emerging challenge for law enforcement. *FBI Law Enforcement Bulletin* **65**(12).

(1997) CDROM vs. Optical Disks. *International Journal of Forensic Computing*, **August**.

(1997) Cluster analysis. *International Journal of Forensic Computing* **January**.

Computer Crime Investigation and Forensics Bibliography. Online documents concerning the investigation of computer crime. http://www.aracnet.com/~gtr/archive/investigate.html.

(1997) Floppy Disks – not a problem. *International Journal of Forensic Computing*, March, Issue 3

Hafner K and Moukoff J (1995) *Cyberpunk: Outlaws and Hackers on the Computer Frontier*.

Henseler J (1997) Developments in forensic computer science. Searching for the digital truth. *International Journal of Forensic Computing*, **September**.

(1997) Investigating floppy disks … *International Journal of Forensic Computing*, **March**.

Schneier B (1995) *Applied Cryptography: Protocols, Algorithms, and Source Code in C*.

Slatalla M and Quittner J (1996) *Masters of Deception: The Gang that Ruled Cyberspace*.

Sterling B (1993) *The Hacker Crackdown: Law and Disorder on the Electronic Frontier*.

Stol C (1989) *The Cuckoo's Egg: Tracking a Spy Through a Maze of Computer Espionage*.

United Nations Crime Prevention and Criminal Justice Programme. 'International review of criminal policy – United Nations Manual on the prevention and control of computer-related crime'. http://www.ifs.univie.ac.at/~pr2gq1/rev4344.html.

(1997) Unix introduction. *International Journal of Forensic Computing* **June**.

Computers *see* **Computer Crime**. **Electronic Communication and Information**. **Fingerprints (Dactyloscopy)**: Automated Methods, including Criminal Record Administration. **Internet Crime**. **Photography and Digital Imaging**: Overview; Digital Imaging Enhancement.

Contamination *see* **Crime-scene Investigation and Examination**: Contamination.

Counterfeiting *see* **Forgery and Fraud**: Counterfeit Currency.

Credit Cards *see* **Forgery and Fraud**: Payment Cards.

CRIME SCENE INVESTIGATION AND EXAMINATION

Contents

Collection & Chain of Evidence
Contamination
Criminal Analysis
Fingerprints
Major Incident Scene Management
Packaging
Preservation
Recording
Recovery of Human Remains
Scene Analysis and Reconstruction
Suspicious Deaths

Collection and Chain of Evidence

J A Siegel, Michigan State University, East Lansing, MI, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0450

Introduction

One of the major activities by police at a crime scene is to search for items of evidence that have some probative value in the crime at hand. The crime scene, in the broad sense, includes more than the direct location where the crime occurred. It may include other physical locations where evidence has been generated. For example, a person may have been killed in one place, the body dumped in another location and the car that transported the body may have been left in a third. All of these are part of the crime scene and all of these contain evidence which has potential to help solve the crime.

Requirements for Preservation of Evidence

Every country that has a rule of law requires that evidence be preserved in the same condition as when it first appears at the crime scene, allowing for changes that take place during analysis of the evidence by the crime laboratory and normal degradation with time. There are two major reasons for this requirement. First, if the evidence is not properly preserved, it may be rendered unfit for analysis by forensic scientists and may lose whatever value it may have had in helping to solve the crime. This may also mean that there may not be enough left for analysis by the defense. This failure to preserve evidence, sometimes referred to as spoilage, may render the evidence inadmissible at trial.

The second major reason for the preservation requirement is that the court must be assured that the evidence is authentic; that is, it is the same exact piece of evidence that was seized at the crime scene. If the evidence is allowed to spoil, it may lose its identifying characteristics.

Chain of Custody

Besides preservation of evidence against spoilage, there are other means to achieve proper authentication. Chief among these is documentation of the evidence. Part of this process is included in the general procedures of recording the crime scene, but there is also the chain of custody. The chain of custody can be thought of as both a process and a document.

Chain-of-custody process

The process that is used to authenticate evidence starts with identification. When evidence is seized

by a police officer, there is a requirement that the evidence (or in many cases, its container) be labeled with name or initials of the recovering officer, date and time of recovery, etc. There may also be a unique case and item identifier that is affixed to the evidence. These identifiers are also used to track the evidence throughout the police department and become a part of the record of the crime scene.

The second step in the chain of custody process is proper packaging of the evidence. How a particular piece of evidence is packaged depends upon the nature of the item and the need to preserve it. Nonetheless there is a general requirement that all evidence be packaged in tamper-evident containers. Note that this requirement does not mean tamper-proof. There is really no such thing as a tamper-proof container. Any container can be breached. The requirement is that, if a container is opened improperly or without authorization, then the tampering will be evident to people who sealed the container originally. The requirement serves to help ensure that the evidence is authentic – that it is the same evidence that was originally seized and that it is in the same condition throughout.

The next step of the chain of custody occurs when the evidence is transported to the crime laboratory. In most cases, evidence is submitted to the laboratory accompanied by a request for laboratory analysis. This form describes the evidence and the type(s) of analysis requested and contains the time, date and location from which the evidence was recovered, as well as whatever identifiers that the police have associated with the evidence. Forensic science laboratories do not, as a rule, rely upon any of these identifiers for tracking the evidence through the laboratory system because laboratory personnel have no way of knowing the source of these identifiers and to what extent they are unique. Instead, the laboratory will assign its own case and item numbers to each item of evidence. This may be done through physically writing the laboratory numbers on each piece of evidence or container, or by use of bar-code labels. This practice is increasing in popularity, at least in the United States, because it makes evidence-tracking through the laboratory system more convenient. When an analyst takes possession of the evidence, the bar-code is scanned and the data entered into a computer. The location and status of any piece of evidence can thus be instantaneously determined by querying the computer. The important thing is that each item of evidence is identified with a unique number and that all of the items in a given case are tied together by the laboratory case number system.

When an evidence container is opened and the evidence examined by a forensic scientist, that person must label each piece of evidence with the laboratory's case number and item number and affix his or her name or initials and the date and time that the evidence was opened. When the analysis is complete, the evidence must be resealed in a tamper-evident manner. This is often accomplished by the use of evidence tape, which is specially designed to stick tightly and shred when attempts to remove it are made. When any type of tape is used, the analyst should put his or her initials and the date and time across the tape so that part of the writing is on the tape and part is on the package. This ensures that, should the tape be removed, it will be difficult to replace without leaving evidence of its removal.

This practice of police departments and crime laboratories of using two or more numbers to a piece of evidence may make the evidence look confusing with its profusion of numbers, initials, dates and times, but this is necessary to maintain the chain of custody. Each party who handles the evidence can use the appropriate markings to identify the evidence.

Chain-of-custody document

Along with the process that establishes the identify and integrity of the evidence throughout its journey from the crime scene to the courtroom, there is also a chain-of-custody document that accompanies the evidence. This is a form that describes the evidence, including the submitting agency and unique case identifiers. It also includes places where one signs and dates the document, signifying possession and control of the evidence. Every time the evidence changes hands, there is a set of signatures of the person giving up custody and the person accepting custody. With this record, any official can instantly determine who had custody of this evidence and during what period of time. The chain-of-custody document accompanies the evidence throughout its journey through the criminal justice system.

The US Constitution and other similar documents throughout the world provide that a person accused of a crime has the right to confront witnesses. Anyone who has come into possession of evidence is a witness in the broad sense and may be called to court to testify as to the fate of the evidence while in his or her possession. This means that a property clerk or intake officer in a crime laboratory who accepts evidence and then passes it on to a forensic scientist, without so much as opening up the container, may be called to court to testify. The chain-of-custody document provides the record that court officials can use to compel an appearance in court from anyone on the chain of custody.

Some police and crime laboratory systems use a separate paper for the chain of custody. This has

multiple copies so that each time the evidence changes hands a copy of the chain may be given to the person who last had custody of it. An increasingly popular method of developing a chain-of-custody document is to incorporate it on to an evidence container. Many evidence bags today come with a chain-of-custody form imprinted on the front. In the case of plastic bags, the part of the bag containing the form is made of a 'writeable' plastic. In other cases the form may have an adhesive backing that can be used to affix the form to any container. This incorporated form has the advantage of accompanying the evidence automatically, without affixing another form to it. Xerox copies can be made of the form each time the evidence changes hands so as to provide a receipt for those who handled the evidence.

Evidence Collection

The saying 'garbage in, garbage out' has often been applied to computer programs, but should also be kept in mind when considering the collection of evidence from a crime scene. The crime laboratory can only analyze what it receives. Evidence that is not protected from spoilage or adulteration may not be in a condition to provide useful information. Also, the problem of insufficient material for analysis can be a major problem in crime laboratories. In many cases, evidence is pretty useless unless there is something to compare it against, so the proper collection of known standards or exemplars is very important. Each of these points is considered in more detail below.

Proper protection of evidence

The methods used to protect evidence vary with the evidence type. Obviously, evidence that is to be searched for fingerprints must not be handled with bare hands, and extreme caution must be exercised in handling it and transporting it to the laboratory. One often sees television detectives picking up a gun by shoving a pencil into the barrel and lifting it, to avoid handling it with bare hands. This practice may alter the markings on the inside of the barrel, making test-fired bullet comparisons with bullets obtained from the crime scene difficult or impossible.

Blood evidence and other human tissue must be handled with care. If fresh or wet blood or bloody garments are packaged in an airtight container, the risk of putrefaction is great and the evidence may be spoiled. In addition bloody garments must be packaged separately to avoid crosscontamination. When handling human evidence, one must also be concerned about transmission of blood-borne diseases such as hepatitis or AIDS.

The crosscontamination problem may apply to many types of evidence. Whenever possible, different items of evidence should be packaged separately to avoid the chance of crosscontamination or the exchange of evidentiary material.

Trace evidence calls for special collection methods. Because there may be very little of this evidence and it may be difficult to see, it can easily be lost. Specially equipped vacuum cleaners containing changeable filters can be used to search for trace evidence over a large area. For smaller areas, trace evidence can be collected by tape lifting. Care must be exercised with tape lifting of certain types of evidence. For example, if paint chips are affixed to tape, the muselage in the tape may adhere to the paint, thus contaminating it. Garments may be searched for trace evidence by brushing them in a clean room on to a large paper surface or into a large funnel.

Sample sufficiency

When it comes to direct, crime scene evidence, one must make do with what is at hand. More cannot be manufactured. If a scene item is limited in size, the crime laboratory will have to live with it and all crime laboratories have provisions for the analysis of sample-limited items. Often, however, insufficient samples are collected by crime scene investigators because of a lack of proper training. This training may be partially the responsibility of forensic scientists, who need to make sure that investigators are aware of the problems that arise with insufficient samples.

Known samples (exemplars)

The collection of exemplars presents its own set of problems. The aforementioned problem of insufficient samples should not be a consideration here. The criminal investigator should be able to collect more than enough material to satisfy the forensic examiners. This may also be a matter of training, so that the investigator is taught to know how much is enough and where to collect exemplars. For example, handwriting exemplars must be collected in certain ways, making sure that, to the extent possible, the writing implements and conditions of collection mimic those of the questioned document. Human head hair exemplars must be collected by combing, not pulling or cutting, from representative areas of the head, and as many as practical must be gathered. Test firings of weapons should be done with the same type of ammunition as the questioned bullets, etc. Each type of evidence has its own peculiar requirements for exemplar collection. Evidence collectors need to be aware of these in order to maximize the chances of successful examination of the evidence.

Negative controls

For certain types of evidence, the presence of proper negative controls is critical to successful interpretation of the evidence. For example, suppose that some suspected blood is found on a carpet near the victim of a homicide. Certainly the stained carpeting would be sent to the crime laboratory for analysis of the blood, but that is not sufficient. Some of the carpeting that has no stain on it, and which is in close proximity to the stained area, must also be collected. It is necessary for the forensic scientist to make sure that there is nothing in the carpeting that could affect the results of the analysis of the blood. The only way this can be determined is by having unstained carpet available for comparison.

See also: **Crime-scene Investigation and Examination**: Recording; Preservation; Major Incident Scene Management.

Further Reading

Osterburg JW and Ward RH (1998) *Criminal Investigation: A Method for Reconstructing the Past*. Chicago: Anderson.

Saferstein R (1998) *Criminalistics: An Introduction to Forensic Science*. Upper Saddle River, NJ: Prentice Hall.

Swanson CR, Chamelin NC and Territo L (1999) *Criminal Investigation*. New York: McGraw-Hill.

Contamination

H B Baldwin, Forensic Enterprises Inc., Orland Park, IL, USA
C Puskarich May, Criminal Justice Institute, Little Rock, AR, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0454

Introduction

The issue of contamination of physical evidence has painfully brought notoriety to several criminal cases. *Webster's Dictionary* defines contamination as: 'to make impure, corrupt, by contact; pollute, taint'. Potential contamination of physical evidence can occur at the crime scene, during the packaging, collection and transportation of the evidence to a secured facility or laboratory, and during evidence analysis and storage.

While forensic scientists in the laboratory are sensitive to the issue and have developed protocols to identify and reduce the risk of contamination, law enforcement has been slower to incorporate precautions in contamination prevention. Recent advances in forensic DNA technology are making it even more important that crime scene personnel become more sensitive to these issues.

The Crime Scene

Crime scene contamination is usually the result of the actions of the personnel at the scene. In general, the greater the number of personnel at the scene, the more likely it is that the scene/evidence will be contaminated. They can deposit hairs, fibers or trace material from their clothing, or destroy latent footwear prints or fingerprints. Footwear patterns can also be deposited by crime scene personnel or anyone entering the scene. As Professor Locard has taught us, when two objects come in contact with each other they exchange trace evidence. Each time we enter a crime scene, we not only potentially leave trace evidence behind, but also take evidence away from the scene.

Forensic DNA analysis has become an increasingly powerful investigative tool. Analysis of biological fluids, and now cells, found at crime scenes can, with relatively high confidence, exclude/include a possible suspect and provide a numerical estimate of the similarity between DNA found at the crime scene and the suspect's DNA. DNA technology now being used in crime laboratories can take samples that are very small or degraded and duplicate and amplify the DNA present to provide a large enough sample for analysis. Because of the analyst's ability to duplicate and amplify very small amounts of DNA from biological evidence, reducing the potential for contamination at crime scenes becomes ever more significant. Single hairs, perspiration and/or saliva inadvertently deposited by an investigator while at a crime scene can now cost valuable time and create the potential for excluding a viable suspect, as well as clouding or confusing the interpretation of the physical evidence.

The level of contamination risk to be expected is related to the type of crime scene and the corresponding number of individuals who have access to it. At a burglary scene, the victim and the officer taking the report may be the only individuals present. In contrast, a typical death scene would usually be visited by the first responder, paramedics, investigators, crime scene examiners, forensic scientists, coroner or medical examiners, prosecuting attorneys and possibly supervisors. Family, friends and neighbors of the victim may also be present. Obviously, owing to the

higher number of individuals in contact with the scene, the potential for contamination would be significantly higher at the scene of a death.

Environmental conditions may also play a major role in the contamination of crime scene evidence. Wind, sun, rain, snow and temperature can play key roles in the destruction of the evidence. For instance, if there is blood at an outdoor crime scene and it rains, the blood may become so diluted that testing it becomes impossible. The same would apply if the blood were exposed to the sun on an extremely hot and humid day. The fluid would be decomposed or contaminated by bacteria to a point where further analysis would be impossible, or inconclusive at best. Wind and temperature are also possible mechanisms for contamination. Wind can carry in contaminates or literally blow evidence away. Temperature in the extremes can obviously cause problems if the items containing evidence become 'cooked' or 'frozen'. This applies to outdoor scenes that are unprotected, but can also apply to indoor scenes with no heating or cooling capabilities. Animal predation may also be a factor for consideration in outdoor scenes.

Presecured scenes

The potential for evidence (or crime scene) contamination increases as the number of people entering a crime scene also increases. Once a scene has been secured, the risk of contamination is greatly reduced. However, what about the events that occurred before the scene was secured? How many people entered and left the scene without being recorded as being there? For example, in a private residence, the victim, the family and the investigating officers are usually the only individuals who have been in the crime scene. In sharp contrast, if a bank robbery occurs, physical evidence such as fingerprints, footwear and trace (as well as biological) evidence could be present from many individuals who recently visited the bank or those customers who were there at the time of the event. A similar situation would also arise if the crime occurred in a motel room or a public park. The amount of potential destruction and contamination of physical evidence in public places such as these presents a real challenge for law enforcement.

Postsecured scenes

Once the scene has been secured, the potential for contamination still exists. The scene is usually secured by only one officer. This is problematic if no provisions are made for others entering the crime scene from other avenues. It is extremely difficult in some situations to thoroughly protect the scene from unauthorized personnel. Once again, in a residential burglary, the scene is less complex and can be secured fairly quickly with minimal personnel. A public park, on the other hand, may be extremely difficult, if not impossible, to secure totally and keep secured until the crime scene is thoroughly processed.

The risk of contamination in all crime scenes is reduced by thoroughly protecting the scene. Consequently, determining the dimensionality of the scene should be the first priority. We cannot protect something we do not recognize as part of the scene. Indoor scenes, by virtue of being enclosed structures, seem easier to secure. Outdoor scenes, on the other hand, are more difficult to secure because of the potential contamination by agents such as weather conditions and crowds. As a result, these types of scene require more personnel for proper protection. Barrier tape, usually yellow or blue and marked by the words 'crime scene' or 'police line', with the additional words of 'do not cross', is used to identify the outer perimeter of the scene. Physical barriers are always needed to define the areas restricted to the public as well as other law enforcement personnel. The barrier used can be as simple as a rope or added markings with attached signs. Providing visual boundaries to the scene assists in restricting access and reducing contamination risks.

Once the area is defined, a command post or rendezvous (RV) point should be established. Forming a command post reduces the potential for contamination of the scene by limiting access of personnel to the scene and identifying who is entering and leaving it. A major crime may require paramedics, firemen, family, friends, neighbors, patrol officers, investigators, supervisors, crime scene personnel, forensic scientists and medical examiners or coroners to be present at the scene. Each person has the potential to destroy or contaminate the scene through normal hair loss and the transfer of fibers and trace evidence from his or her own environment (home, office or vehicle) to the crime scene. Even the family pet can contaminate the scene by the introduction of additional trace evidence and the transfer of evidence from one area of the scene to another.

Equipment

The equipment used in documenting and processing crime scenes also represents a possible source of contamination. Crime scene personnel need to be aware of the possible crosscontamination that can be caused by their equipment. As crime scene examiners travel from one scene to another they run the risk of transferring hairs, fibers and biological fluid from other scenes to the new crime scene. This contamination can be easily controlled if crime scene personnel

decontaminate their equipment before and after attendance at each crime scene. Equipment that should be decontaminated includes, but is not limited to, clothing, note pads, photography equipment, sketching equipment and all processing equipment in crime scene kits.

Crimes involving more than one scene have similar contamination issues, particularly if a suspect, or more than one suspect, commits a series of crimes in the same time frame. The situation in Littleton, Colorado, USA, for example, where two suspects shot several people in various locations in the huge school complex, could have posed a severe contamination issue for law enforcement. If multiple scenes are involved, physical evidence can be easily transferred from one scene to another if precautionary procedures are not applied by the crime scene personnel. Therefore, when multiple scenes are involved, the processor(s) must be aware of crosscontamination issues and take steps to reduce their occurrence. The involuntary transfer of evidence from one scene to another must be taken into consideration and addressed by the law enforcement personnel at the scene. This can be easily addressed by identifying requirements for personal protective equipment (PPE) and a decontamination zone, or even utilizing different staff for scene and suspect.

Required PPE consists of a mask, jumpsuit, gloves, booties and head cover. All of these items must be disposable. Most crime scene personnel wear booties, a mask and gloves. This is usually done as a biohazard exposure precaution rather than for reducing contamination risk. However, the use of PPE is also an effective mechanism for reducing contamination potential and subsequently increasing the investigative value of biological evidence which may be subjected to forensic DNA analysis. In some jurisdictions, PPE is exhibited to prove contamination issues were addressed in the investigation.

To reduce the potential for crosscontamination, a decontamination zone must be established. This safe zone is the area where crime scene equipment and PPE can be safely cleaned, removed and properly discarded. For some scenes, the decontamination zone may be more elaborate than in others. For instance, at a burglary scene, the decontamination zone is simply a ‘safe’ area of the residence where crime scene equipment can be placed. By having equipment and supplies restricted to one area, the potential for contaminating other evidence from the scene is minimized. In a violent scene or any scene with abundant biological fluids, crime scene personnel must be careful not to transfer biological fluid from their footwear to different parts of the scene. Biological fluids also create a safety issue: therefore, crime scene personnel should be required to wear the full array of PPE at these types of scenes.

The decontamination zone must contain the appropriate cleaning supplies and all equipment for the decontamination of the person as well as the equipment. Usually, decontamination involves the removal and discarding of disposable clothing and the wiping down of all equipment with a 10% solution of bleach. Other disinfectants may be appropriate for different equipment, but the most convenient as well as the ‘standard’ of the safety industry is a 10% solution of bleach in water. The use of bleach will require that a plastic sheet or, in cases where the PPE needs to be decontaminated, a small plastic child’s pool (about 1 meter in diameter) be used. This would not only protect the floor-covering from the bleach, but also offer the ‘best’ way to safely decontaminate PPE when crime scene personnel move from one scene to another in multiple scene crimes.

Even after the crime scene is well secured, contamination risks may still be high. Not only do crime scene personnel frequently walk through the scene, but their equipment has the potential of transferring unassociated trace items in the scene as potential evidence. Furthermore, the processing techniques used by crime scene personnel to find or enhance various forms of evidence can also contaminate the evidence. Typically, the crime scene can yield a wealth of potential evidence such as fingerprints, footwear prints, toolmarks, hairs, trace material, biological fluids and latent patterns. Each type of evidence may require the use of chemicals or powders to identify or enhance them. To reduce the risk of contamination, crime scene personnel should follow a protocol for evidence collection that inhibits evidence destruction and contamination. An order of evidence collection that would reduce contamination is: recover the trace, hairs and fibers first; then biological fluid; tool marks; visible fingerprints or footwear patterns; then, lastly, latent patterns that require powder or chemical enhancement.

Even then, the decision on what processing steps are needed in the crime scene is left to the judgment of crime scene personnel, based on the evidential potential of that item. A pry bar left at a crime scene by the suspect can illustrate this point. A pry bar was used to pry open a door or window to gain entry into a structure. The evidence potential of that item alone includes latent prints, biological and trace material, transfer evidence, tool marks and control standards. Crime scene personnel must keep the potential of the evidence in mind, but must not forget the practicality of the evidence. Physical evidence can support the facts of the case, identify a crime was committed, identify the suspect or victim, develop

modus operandi, and prove or disprove the theory of the case.

Packaging and Collection

Evidence is packaged to prevent destruction and contamination. New containers must be used to package all evidence. For some evidence, sterile containers are required. The packaging equipment must also be free of contaminates. This is easily accomplished by keeping all the packaging supplies in one case and handling them carefully from one scene to another.

Evidence containers must be sealed at the crime scene. This reduces the contamination potential and keeps the integrity of the evidence intact. When containers are left open and removed from the scene, the potential for contamination increases dramatically. Consequently, containers must be properly sealed and marked for identification at the crime scene.

Because of the sensitivity of the forensic DNA analyses currently being performed by crime laboratories, handling biological evidence properly is critical. Drying items with biological fluids on them prevents destruction or contamination by bacteria. Drying wet objects at the scene may be impossible and, at times, not appropriate. It is recommended that wet items be packaged in a paper container and sealed, then placed in an open plastic container. The plastic container is used only as a transportation tool. It prevents biological fluid from crosscontaminating other paper containers of evidence and stops the fluid from leaking through the paper packaging on to the floor or seat of the person's vehicle. Once removed from the scene and transported to the police or laboratory facility, wet items can then be removed from the plastic container and dried in a sterile vented hood. The wet evidence can be placed in the drying hood with the paper container still sealed. It will dry without being removed from the container. If excessively wet, the item will need to be taken out of the container, dried and repackaged. The original container must be kept to maintain chain-of-custody and evidence integrity. Any trace evidence that may have fallen off the item must be retained in the paper container. If a bloody item is to be dried in a vented hood, the hood should be decontaminated first. In addition, items from different cases must not be placed in the hood at the same time. Access to and from the room where the hood is located should be restricted and monitored for security purposes.

Transporting Evidence

Transportation of the evidence from the scene requires special consideration to prevent destruction and contamination. Some evidence, particularly biological evidence, may be sensitive to absolute temperature or fluctuations in temperature. During the summer months, or when a vehicle is in direct sunlight, the temperature inside a vehicle can rise substantially. Placing evidence in an enclosed area of a vehicle under these types of circumstance can destroy or contaminate it. During colder times of the year, precautions to prevent freezing of the evidence should also be taken.

To prevent contamination from the crime scene person's perspiration falling on to the packaging and then on to the evidence, packages must always be handled while wearing disposable gloves. Leakage from the evidence containers can be contained by transporting the evidence in an open-top plastic container.

Special precautions must be taken when transporting the evidence from the scene to a temporary facility, such as the crime scene person's office. Provisions must be made to store the evidence properly to prevent contamination and maintain evidence security. The ability to store the evidence temporarily in a secured facility, away from other items, in a temperature controlled environment, must exist or the evidence must be transported to the crime laboratory immediately. Access to these areas at the weekend or after working hours must also be controlled and monitored for evidence security purposes.

Crime Laboratory Analysis

To submit evidence for analysis at a forensic laboratory, the evidence is taken to a 'sign in' area of the facility, where all evidence is received. This is another area for potential contamination. Evidence from other cases can have a leakage problem and consequently contaminate all evidence packages placed on the receiving counter. Decontamination of this area should be done on a repeated basis during the working hours of the laboratory. After the evidence is properly received, it usually goes to a temporary storage vault. Potential leakage of other containers in this vault, or the storage area in general, may cause contamination problems that need to be addressed. Eventually the evidence is removed from the vault and taken to a section of the laboratory for examination and analysis. The package is now placed on a table or counter where many other pieces of evidence have been over the years. However, most laboratories have adequate decontamination procedures already in place. The analysts are cognizant of crosscontamination issues and keep their work areas decontaminated on a routine basis. Standard procedures and policies are usually adopted by the facility and the

forensic scientists to reduce the potential risk of contamination.

After the Analysis

After the evidence has been analyzed, the evidence containers are usually resealed in the same packaging or an additional package and stored in a temporary evidence vault. Eventually it is transferred to the investigating agency or retained by the laboratory for court. Transporting the evidence to storage and additional handling may also create potential contamination concerns and must not be taken lightly. Further analysis of the evidence may be required at a later date, sometimes years later. Paper containers have the potential of allowing passage of fluid through their walls because they are 'breathable' and porous containers. Protective gloves should therefore always be worn when a paper container is handled.

Conclusion

The potential for evidence contamination has been a concern of law enforcement and forensic practitioners, in general, ever since evidence was first analyzed. However, the potential impact of contamination of evidence upon the outcome of a criminal investigation has become ever more important, owing to the sensitivities of current scientific analyses such as forensic DNA analysis. If evidence is properly collected from the scene, packaged and handled correctly during transportation and storage, and decontamination procedures are used, the potential for contamination will be greatly reduced. As a result, the integrity and value of the evidence will be maintained, regardless of what additional analyses are developed in the future. Continuing education and training for law enforcement and forensic specialists is required to insure the proper handling of evidence from scene to storage, ultimately reducing the risk of contamination as well as the impact of these issues upon the outcome of a criminal investigation.

See also: **Crime-scene Investigation and Examination**: Recovery of Human Remains; Preservation; Contamination; Major Incident Scene Management.

Further Reading

Eliopulos LN (1993) *Death Investigator's Handbook: A Guide to Crime Scene Processing, Forensic Evaluations and Investigative Techniques*. Boulder, CO: Paladin Press.

Fisher BAJ (2000) *Techniques of Crime Scene Investigation*, 6th edn. Boca Raton, FL: CRC Press.

Hawthorne MR (1999) *First Unit Responder: A Guide to Physical Evidence Collection for Patrol Officers*. Boca Raton, FL: CRC Press.

US Department of Justice, National Commission on the Future of DNA Evidence (1999) *What Every Law Enforcement Officer Should Know About DNA Evidence*. Washington, DC: Department of Justice.

Criminal Analysis

O Ribaux and **P Margot**, University of Lausanne, Lausanne-Dorigny, Switzerland

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0462

Introduction

Traditionally, each scene of crime is seen as an independent 'complete' set of data from which the intelligent investigator can find the signature of the offender. Links between cases are often inferred through police investigation; in a separate process, cases are also compared through the use of physical evidence collected at the scene of crime. The question of how to integrate the different pieces of data available into a coherent framework for intelligence purposes has become an important topic of research.

Increased mobility and new communication channels give criminals the capacity to better plan and organize their activities over large geographical areas. Patterns reflecting fraudulent activities are therefore always more difficult to reveal from the huge quantity of data scattered in files across independent police and legal structures.

A key part of law enforcement is to understand those activities, through the development and use of methods, models and tools for collecting and then interpreting the large volume of data available in real time. Consequently, intelligence programs have been developed, leading to the recognition of a field of activity called crime analysis, which has been described as 'the identification of and the provision of insight into the relationship between crime data and other potentially relevant data with a view to police and judicial practice'.

From an intelligence perspective, forensic science plays an important role in crime analysis, in that forensic science covers a broad range of methods and techniques aimed at revealing relationships between people, scenes of crimes and objects, as well as helping to develop working hypotheses. The great potential of physical evidence for linking crimes and

forensic scientists to reduce the potential risk of contamination.

After the Analysis

After the evidence has been analyzed, the evidence containers are usually resealed in the same packaging or an additional package and stored in a temporary evidence vault. Eventually it is transferred to the investigating agency or retained by the laboratory for court. Transporting the evidence to storage and additional handling may also create potential contamination concerns and must not be taken lightly. Further analysis of the evidence may be required at a later date, sometimes years later. Paper containers have the potential of allowing passage of fluid through their walls because they are 'breathable' and porous containers. Protective gloves should therefore always be worn when a paper container is handled.

Conclusion

The potential for evidence contamination has been a concern of law enforcement and forensic practitioners, in general, ever since evidence was first analyzed. However, the potential impact of contamination of evidence upon the outcome of a criminal investigation has become ever more important, owing to the sensitivities of current scientific analyses such as forensic DNA analysis. If evidence is properly collected from the scene, packaged and handled correctly during transportation and storage, and decontamination procedures are used, the potential for contamination will be greatly reduced. As a result, the integrity and value of the evidence will be maintained, regardless of what additional analyses are developed in the future. Continuing education and training for law enforcement and forensic specialists is required to insure the proper handling of evidence from scene to storage, ultimately reducing the risk of contamination as well as the impact of these issues upon the outcome of a criminal investigation.

See also: **Crime-scene Investigation and Examination**: Recovery of Human Remains; Preservation; Contamination; Major Incident Scene Management.

Further Reading

Eliopulos LN (1993) *Death Investigator's Handbook: A Guide to Crime Scene Processing, Forensic Evaluations and Investigative Techniques*. Boulder, CO: Paladin Press.

Fisher BAJ (2000) *Techniques of Crime Scene Investigation*, 6th edn. Boca Raton, FL: CRC Press.

Hawthorne MR (1999) *First Unit Responder: A Guide to Physical Evidence Collection for Patrol Officers*. Boca Raton, FL: CRC Press.

US Department of Justice, National Commission on the Future of DNA Evidence (1999) *What Every Law Enforcement Officer Should Know About DNA Evidence*. Washington, DC: Department of Justice.

Criminal Analysis

O Ribaux and **P Margot**, University of Lausanne, Lausanne-Dorigny, Switzerland

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0462

Introduction

Traditionally, each scene of crime is seen as an independent 'complete' set of data from which the intelligent investigator can find the signature of the offender. Links between cases are often inferred through police investigation; in a separate process, cases are also compared through the use of physical evidence collected at the scene of crime. The question of how to integrate the different pieces of data available into a coherent framework for intelligence purposes has become an important topic of research.

Increased mobility and new communication channels give criminals the capacity to better plan and organize their activities over large geographical areas. Patterns reflecting fraudulent activities are therefore always more difficult to reveal from the huge quantity of data scattered in files across independent police and legal structures.

A key part of law enforcement is to understand those activities, through the development and use of methods, models and tools for collecting and then interpreting the large volume of data available in real time. Consequently, intelligence programs have been developed, leading to the recognition of a field of activity called crime analysis, which has been described as 'the identification of and the provision of insight into the relationship between crime data and other potentially relevant data with a view to police and judicial practice'.

From an intelligence perspective, forensic science plays an important role in crime analysis, in that forensic science covers a broad range of methods and techniques aimed at revealing relationships between people, scenes of crimes and objects, as well as helping to develop working hypotheses. The great potential of physical evidence for linking crimes and

bringing together other solid informative data has been shown, specifically through the use of multiple databases. Such developments have also greatly influenced crime scene management, because collection and treatment of data are highly interlinked.

An overview of crime analysis is given in the first part of this article. Despite the variety of independent developments, generally tied to the type of trace under consideration (fingermarks, shoemarks, biological markers, etc.), databases in forensic intelligence have a generic flavor. For instance, linking or pointing to a suspect already known for antecedent activities from the data recorded are common facilities provided by some of those systems. Classification through a description of *what* the systems are helping to do in terms of elementary inferences will place the current role of forensic intelligence within crime analysis.

Finally, both ends will be joined through the example of an integrated approach for managing burglary using geographic/temporal data, modus operandi and physical evidence.

Crime Analysis

The context

Developments in the field of information technology provide the police with a vast quantity of data, from different sources, reflecting criminal events and their relationships. Relevant patterns are not easy to extract from this huge quantity of noisy data. Moreover, the available data are often incomplete and uncertain; for instance, all criminal events are not reported to the police; the date and time of a particular event are generally not precisely known; and traces collected are often fragmentary. Dealing with imperfect data reflecting complex evolving phenomena is very common to criminal investigation and crime analysis.

Legal and organizational barriers, the proliferation of methods for collecting and interpreting data, and the consequent variety of incompatible recording systems make comparisons difficult, when performed. Poor analysis of similar offences crossing different police areas is recognized as one of the major weaknesses pertaining to police information systems; this has been called 'linkage blindness'.

Finally, the analysis of data is generally distributed between people with different knowledge, experience and culture. For instance, cooperation between the laboratory scientist, the crime investigator and the crime analyst is of increasing importance to provide the best use not solely of new techniques but also of scientific attitudes during the whole investigation.

As a consequence police have adapted information systems:

- New structures have been created to favor exchange of information across countries.
- New intelligence structures within the organizations have been created, with an important part dedicated to crime analysis.
- Structured and normalized methods of analyzing data have been developed.
- A broad variety of computerized tools have been introduced (geographic information systems, drawing programs, statistical packages, broad variety of databases, etc.).

Brief historical perspective

The use of methods of collecting and exploiting large quantities of data for intelligence purposes is not new. At the beginning of the twentieth century, a number of authors described ways of recording data in filing systems; fingerprints are a well-known example, as is Bertillon's anthropometry. The proliferation of methods, as well as the development of a broad variety of incompatible classification systems, brought a response from Locard in 1906; he was already arguing for the creation of a uniform international method of recording data that could help identify people independently of borders.

During that same period, Reiss, in 1914, attending the congress of Monaco, and Locard not only recommended the harmonization of methods but also emphasized the advantages of using the different types of data available. These important works, pertaining to recording methods, have been reviewed, adapted and their efficiency increased through the use of computer technology. Simultaneously, the evolution of organized criminality highlighted the need to improve police information systems. Intelligence processes have been defined and crime analysis was recognized as a main component.

The origin of crime analysis is attributed to North America in the 1960s. In 1994, an expert group published the *Crime Analysis Booklet*, now a recognized reference book in police agencies in Europe: general forms of data analysis are defined; how to set up a crime analysis function within an organization is also considered.

Goals of crime analysis

Roughly, crime analysis aims at:

- deciphering the knowledge used by experienced investigators to identify and formalize concepts and notions, and the methods used by them to marshal their information;
- conceiving new methods of treating data, and adapting existing ones, given that computer tools can improve the analysis process in a way that was

not previously possible, and that criminality itself is evolving over time, and that weaknesses observed in the course of an analysis can necessitate upgrades;
- normalizing language and symbols used in the course of the analysis in order to facilitate teamwork on complex problems, and to disseminate the results of the analysis in a way that is easy to interpret;
- using those methods to produce conclusions/hypotheses that can help towards concrete actions.

A broad variety of analysis forms have been defined; for instance, crime pattern analysis, profile analysis, case analysis (course of events immediately before, during and after a serious offence), comparative case analysis, etc. Forensic science data are exploited in most of these different forms of analysis, as will be shown in the rest of this article.

Databases in Forensic Intelligence

Forensic intelligence is a process that starts from the collection of data, often at the scene of crime. The exploitation of data relies entirely upon the quality and quantity of the collected data. It could therefore be said that inferences drawn, and action taken at the scene, are fundamental to forensic intelligence as they will entirely determine what data will be collected and then exploited. That part of the process will not fully enter into consideration in this article, as these methods are extensively described in other parts of the encyclopedia.

Computer systems in forensic intelligence can be classified into two groups. Locard, in 1920, noticed that investigative methods are essentially based on analogy. Consequently, it is not surprising that most existing computer systems recognized as forensic intelligence systems can be viewed as an aid to performing this type of reasoning process.

The second class of system pertains to collections that help classify evidence by the type of object that could have transferred a trace.

Analogies

Similarities found between accessible information can lead to inferences on the profile of offenders, their mode of action and the means they have used. The identification of recidivists is the better known form of this basic scheme. From a new situation or a new case, the purpose is to identify known offenders from their antecedents. The search for links between cases is another elementary activity that falls within the same framework; it allows groupings that delineate criminal events. Finally, when objects are found or perpetrators identified, all the cases in which their participation or presence can be supposed should be extracted. This process is traditionally incomplete because links may not systematically be searched *before* the offender has been arrested. It could be that identifying an offender for a crime is sufficient to put him or her behind bars, without being charged with other offences that could be difficult to prove. This is something that real-time crime analysis avoids.

These steps have been partially reproduced in computer applications; the best-known illustrations are AFIS (automatic fingerprint identification systems) and DNA databases. Thus, if a fingermark, even fragmentary, found at the scene of a crime is compared with the content of a database, a restricted number of similar fingerprints will be given to the experts who interpret them. Potentially, all the other marks transferred, or the information left by the offender (from clothing or accessories), could be analyzed in a similar way.

Differences and difficulties The forensic intelligence process starts with the collection of data and ends with the integration of results into the analysis of crimes under investigation. If we assume that at least one database has been developed in this process, important intermediate steps may be described as follows (**Fig. 1**):

- the acquisition of new data in order to transform it into a digital or symbolic form suitable for searching for similar items in the database, and to memorize it in a way that will allow retrieval when needed;
- the search and matching process;
- the interpretation/verification of the returned result.

The people participating in this process for each database may range from the scene of crime officer to the expert responsible for the interpretation step. To ensure that the process is reliable, and to diminish time and resources needed, three general rules may be applied:

- check existing data, their coherence and their 'usefulness';
- write computer programs that implement part of the process, where possible;
- where it is not possible, define methods and control their application.

The specific difficulties pertaining to the nature of data have led to the separation of the implementation of the process for each type of information. Biological traces, fingermarks, shoemarks, bullets, cartridges and toolmarks are generally independently treated

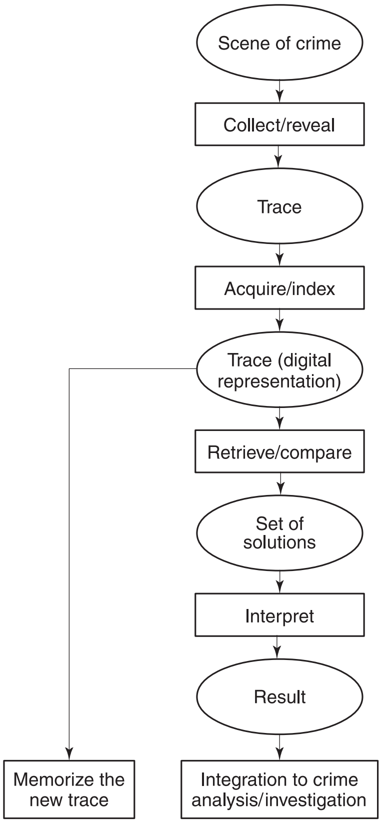


Figure 1 Forensic intelligence process, from the collection of data to the dissemination of results.

in separate databases. What follows is not an exhaustive inventory of existing systems and the way they function; it is only aimed at suggesting a framework for evaluating difficulties related to the computerization of systems pertaining to specific types of trace. The collection phase is described elsewhere.

The acquisition problem The acquisition problem is the problem of transforming data (trace) into digital or symbolic form suitable for comparison with the content of the database, and then memorizing it. This operation can be long and complex, and sometimes not fully controllable by machine. For instance, from a DNA point of view, it includes the application of the analytical techniques used to identify markers of a sample, together with encoding into the system the obtained values in the form of numbers.

Most information is gathered through a device that is interfaced with a computer, like a scanner or various types of camera. The image obtained is then treated to extract relevant features for encoding. This operation is computerized to different degrees, depending on the trace and its quality.

Full automatic processing of good quality 10-print cards is sometimes possible, but computer workstations are generally designed with image enhancement processes, and interactive graphics editors mark minutiae and other features. Poor quality traces regularly require this manual operation.

Databases recording shoemarks provide an example that could be situated, in our framework, between DNA and fingerprints. Encoding by hand is difficult owing to the variety of, and changes in, sole patterns. Such systems give good results, but difficulties are encountered in maintaining the quality of data when there is more than one operator. Automatically extracting patterns from a fragmentary shoemark, or when the mark is concealed under irrelevant data, is now a research topic but is computationally very complex, if not intractable for poor marks. The best practical solution is to capture the image through a device and to provide a graphic icon editor that helps a human operator to encode the mark.

Retrieval and matching Collection of data at the scene of crime is always incomplete and imprecise, and collected marks are often fragmentary, even if the investigation is careful and thorough. An object that caused a trace can evolve, and marks or prints can be distorted. The match between recorded data and collected evidence (in its digital form) is therefore generally only partial. A human operator must always interpret a limited set of possible solutions at the end of the chain.

A broad range of retrieval and matching algorithms using various techniques (statistical, neural networked-based or coming from artificial intelligence) are actually implemented to compare the data automatically. They must be designed so that the result of a search is a limited ranked list of possible matches; this avoids painful and time-consuming fruitless comparisons at the end of the process. Those algorithms are generally rarely published owing to economic interests, because they can be viewed as the heart of such databases.

From a practical point of view, DNA and fingerprints provide extreme examples; the problem of matching two DNA markers is very simple from a computing perspective, as the match is generally determined by an 'exact' comparison. (It is not claimed that the implementation of the whole process, from the collection of data to its storage in the system, is easy, but the 'retrieval and matching problems' themselves are not difficult.) A retrieval algorithm can be implemented directly with simple computer development tools. Common databases found on the marketplace are sufficient for this purpose. Comparing fingerprints is a far more complex task owing to the quantity of information to be treated and the imperfect nature of this type of collected data.

There are other fields where such matching pro-

cesses are efficient, like shoemarks, bullets and cartridges; but yet other fields demonstrate difficulties in overcoming the imperfection and complexity behind the type of treated information. Toolmarks, handwriting and voice recognition are three examples.

Classification

One other elementary criminal investigation question where forensic science plays a fundamental role is the determination of the type of object from a trace or evidence collected at the crime scene. For example, in hit-and-run car accidents, it is important to know quickly the make and model of a vehicle. This is possible with paint flakes found at scenes and the maintenance of a systematic database that helps determine correspondence. Knowledge of the type of printer that has been used to produce a document or a counterfeit banknote can constitute an essential element of investigation. Databases of toners can help in this determination.

The difficulties encountered when using this type of scheme not only pertain to the nature of the data treated; market evolution, fashion, etc. render permanent upgrading essential to ensure that memorized classes are adequate and correspond to reality. This requires considerable effort because the field of interest can change rapidly: firearms, types of vehicle, printers, photocopying machines, types of shoe, types of fibers and cosmetics rapidly appear and disappear. Such efforts are generally distributed in collaboration between different forensic laboratories.

Forensic Intelligence and Crime Analysis

Some kinds of problem that forensic science data help to solve may be explained from an intelligence perspective. The very basic inference patterns identified are obviously required in a broad variety of investigations. It may even be said that linking evidence with a suspect, or establishing that the same perpetrator was present at two different scenes, is fundamental, independently of the type of offence.

The combined use of data has also to be considered. For instance, from a methodological point of view, it is sometimes possible to assume a link between two offenses, based on the comparison of shoemarks found at the scenes. These hypotheses could be tested through the comparison of comparable marks, such as toolmarks, collected at both scenes. From a technical perspective, it is also possible to connect databases together to simulate such a process, not only when preparing evidence for the court but also in the investigative part of the process.

Unfortunately, the reality is, for the moment, less exciting. Centralized police databases record forensic science data very poorly, generally in a useless form, and existing forensic intelligence databases are often not designed to be connected with other types of information, not only for commercial or technical reasons, but also because of legal restrictions. (We have seen that one of the most important parts of the problem of retrieving a trace in a database is the matching problem. Generally, police (administrative) databases use technical tools that can only perform 'exact' matches, which are not adequate for the kinds of problems that arise.)

The second central question pertains to knowing whether forensic science data have a broader role to play in crime analysis than in the restricted set of inferences drawn above.

The reasoning processes used in crime investigations are exceedingly complex, and forensic science data can play an important role within other forms of analysis, in particular to help delineate criminal events. This is evident in a subfield of forensic intelligence: drug intelligence. One could hope to better understand the market and organization of drug dealing, the manner used to penetrate a country, and how drugs are distributed and transformed in clandestine laboratories, through the analysis and comparison of seizures. Linking seizures is one basic element of drug intelligence, but the questions pertaining to the significance of links are very difficult to solve. Extensive efforts have been dedicated to the use of analytical techniques in this context, but forensic science currently lacks models to help the systematic management of that information and its dissemination.

The next section will demonstrate how linking through forensic science data is integrated within crime analysis; inference patterns in which forensic science data can play a role are also explained.

Crime Analysis Methods That Use Forensic Science Data

The method

The crime analysis part of the intelligence process for serial crimes can be viewed as an incremental process that accepts new 'cases' as input, and elaborates and maintains a set of relations in the growing set of cases. These memorized relations can provide the basis for representing series or phenomena (**Fig. 2**). At each iteration, the elaboration of new relations must also rely on the examination of the currently known structure of the 'memory', as the picture of the phenomena changes over time and can influence the search for new information.

From this picture, a broad range of operational actions can be taken to prevent subsequent cases,

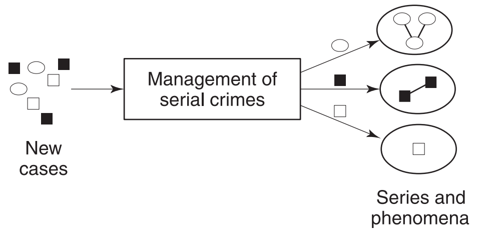


Figure 2 Management of serial crimes can be viewed as a grouping process.

even identify and arrest suspects, or describe the offender's 'profile' (geographic, timing, modus operandi, other habits, etc.). A more strategic point of view is to extract from this knowledge a policy to fight against the type of crime of interest.

Traditionally, such methods focus primarily on personal data, supplemented by modus operandi and geographical considerations, but forensic science data have been proved to be of value in linking cases together. As linking will obviously be one of the basic inferences performed, it can therefore be said that forensic science data can play an important role in the process for different types of serial offence.

Possible models for that process have been developed, using techniques ranging from statistical methods to artificial intelligence-based techniques, some using forensic science data. A concrete example shows how information can be obtained for serial burglary, using only systematic comparisons of shoemarks.

Examples

Example 1: linking house burglaries During the period between November and March (winter) each year, series of burglaries are perpetrated between 1700 and 2200 hours. This phenomenon is probably connected with the ease with which the offender can see if a house is occupied. During the winter, and between these hours, it is dark and very easy to see signs of activity in groups of houses. Moreover, the escape after the crime is not too risky, as the chosen point of entry into the premises is often a back window, which also provides an easy emergency escape route. Another advantage is that there is still quite a lot of activity in the streets, enabling the offender to melt into the background and disappear.

Extracting this phenomenon from all other burglaries, it is possible to compare all the shoe marks collected at the scenes. **Figure 3** shows the picture obtained for one police area (the Swiss canton of

Figure 3 Result of the systematic comparison of shoemarks in the canton of Vaud, during the winter of 1997–1998. Each circle (or triangle) represents a case. White circles indicate that no information was found at the scene, or no linkage was possible on the basis of the collected information. Cases identified by the same symbols (in both shape and tone) have been assumed to belong to the same series. This representation gave an accurate picture of what had happened during the period of time under review.

Vaud). From this knowledge, it is possible to generate the profiles of the different series identified (geographical, modus operandi used, etc.), and take some operational and preventive decisions.

Example 2: tendencies analysis Another analytical orientation in the study of serial burglary is to determine tendencies. For instance, in the same area, it was discovered that the number of burglaries perpetrated during the night, often while the victim was sleeping, increased from April to June. Once identified, this important phenomenon was correlated with the activity of organized teams active on a broad geographical scale. Physical evidence helps strengthen the links or even demonstrates the supposed link.

This tendency has been highlighted through the use of geographical and modus operandi parameters, which is quite common in the practice of such forms of analysis. Forensic science data are also shown to play an important role in this form of analysis, for instance using the width of the toolmarks collected from the scene, independently of all other data, as shown in **Fig. 4**. This is an example of an other inference where forensic evidence can be efficiently used in crime analysis.

This process is efficient in clustering (grouping) cases together on a multiperspective base where forensic science data does effectively complement other police data.

Conclusion

Much work should be dedicated in the near future to the analysis of inferences used in solving the different problems of crime analysis, including forensic science data. Architectures are being researched and built, with techniques for extracting knowledge from collected information. There may be potential for computerization with statistical, neural networked or artificial intelligence techniques (data mining techniques). Forensic scientists play an even more important role in the resolution of these problems, given that they can deal with the specific nature of the collected forensic science data. They provide the scientific attitude needed to approach such problems, with an extension of the role of forensic science, as well as a great potential for crime analysis.

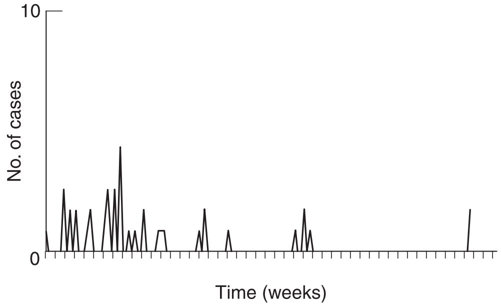


Figure 4 The number of burglary cases in a Swiss canton where the offender has used a tool of 14–16 mm width. The observed tendency has been further studied in order to identify the series. Courtesy of B Sermier, Institut de Police Scientifique et de Criminologie, Université de Lausanne, Switzerland.

See also: **Crime-scene Investigation and Examination**: Recording; Collection and Chain of Evidence; Recovery of Human Remains. **Deoxyribonucleic Acid**: Databanks. **Document Analysis**: Handwriting. **Drugs of Abuse**: Classification, including Commercial Drugs. **Fingerprints (Dactyloscopy)**: Automated Methods, including Criminal Record Administration. **History**: Forensic Sciences. **Pattern Evidence**: Tools. **Voice Analysis**. **Serial Killing**. **Geographical Profiling**.

Further Reading

Crime Analysis Working Group (1997) *Crime Analysis Booklet*, 2nd edn. Lyon: Interpol.

Kind SS (1987) *The Scientific Investigation of Crime*. Harrogate: Forensic Science Service.

Locard E (1920) *L'Enquête Criminelle et les Méthodes Scientifiques*. Paris: Flammarion.

Fingerprints

D L J Clegg, Nottinghamshire Police, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0456

Introduction

The discovery and reporting of a crime will in most cases lead to the start of an investigation by police, who may often rely on the specialized skills of the crime scene investigator to locate, evaluate and record the physical evidence left at, or removed from, the scene. Such investigations are reliant upon the physical transfer of material, whether it be obvious to the eye or otherwise. It is the successful collection of this evidence that often assists the police in the prosecution of an offender before the court. It has been said that 'Few forms of evidence can permit the unquestionable identification of an individual ... and only digital patterns [fingerprints] possess all the necessary qualities for identification'. It is the forensic use of these fingerprints that will now be addressed.

The skin on the inner surface of the hands and soles of the feet is different from skin on other areas of the body: it comprises numerous ridges, which form patterns, particularly on the tips of the fingers and thumbs. Examples of three primary pattern types, arch, loop and whorl, can be seen in **Fig. 1**.

The ridges are not continuous in their flow, in that they randomly end or bifurcate. An example of a ridge ending and a bifurcation can be seen in **Fig. 2**. The ridge endings and bifurcations are often referred to as 'characteristics', 'Galton detail' (named after one of the early and influential pioneers of the science), 'points' or 'minutiae'. The characteristics often combine to create various configurations and these have been assigned names, as can be seen in the examples provided in **Fig. 3**. The 'dot' (Fig. 3D) is often the subject of debate as to whether it is uniquely a characteristic, such as the ridge ending or bifurcation, or a combination of two ridge endings; however, such debate is of academic value only. The ridges themselves provide us with the ability to grip and hold on, and the skin in this context is referred to as friction ridge skin.

On closer examination of the ridges it can be seen there are small sweat pores on their upper surface area and no hairs are present. The perspiration, which is exuded through the sweat pores, is primarily water; on reaching the surface it will be deposited along the tops of the ridges. Additional matter or contaminates may also be present on the ridges, collected from oily regions of the body or some other foreign source, such as food, dirt, grease or blood. When an area of friction ridge skin comes into contact with a surface, a deposit of the perspiration or contaminate may well be left on that surface, leaving an impression of the detail unique to the individual.

A friction ridge skin impression is generally difficult to visualize and is often referred to as a latent deposit. The impression may, however, be visible if it has been deposited in dirt, blood or some other easily seen medium. There are some acids contained in the perspiration which may cause the impression to become 'etched' into some soft metals, while friction ridges may also be impressed into soft media such as putty.

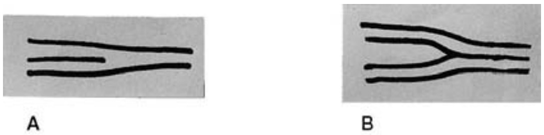


Figure 2 (**A**) Ridge ending. (**B**) Bifurcation.

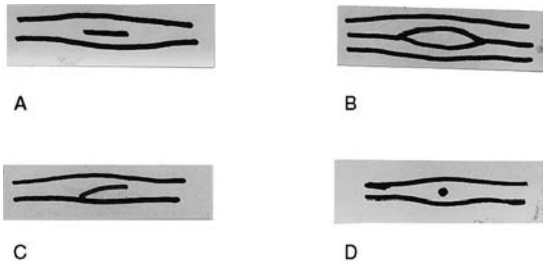


Figure 3 Configurations: (**A**) island or short ridge; (**B**) lake or enclosure; (**C**) spur; (**D**) dot.

Examining the Crime Scene

The crime scene will be examined by a scene of crime officer or a fingerprint technician. In either case it will be the responsibility of that person to conduct an examination of the scene in an effort to locate and develop fingerprints that may have been left by the person who committed the offense. It has been said that a scientific methodology must be applied to the crime scene examination, as follows:

1. Make observations.
2. Arrive at a hypothesis.
3. Test the hypothesis against the physical evidence observed until it cannot be refuted.

Surprisingly, many examiners do completely the opposite when they undertake a crime scene examination.

The examination will generally be started at the point at which entry was gained, if this can be determined. Consideration will be given to all surfaces or objects that have been touched or handled during the commission of the crime, including the point of egress, should this be different than the point of entry. For this examination to be thorough, it will be necessary to speak with either the investigating police or the victim of the crime. Should the crime

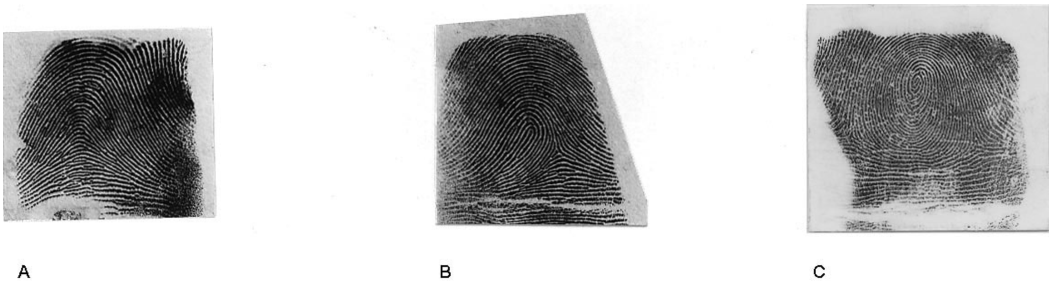


Figure 1 Fingerprint patterns: (**A**) arch; (**B**) loop; (**C**) whorl.

scene officer or fingerprint technician have the opportunity of speaking with witnesses, vital information regarding items from or areas of the scene may be obtained. The opportunity of speaking with witnesses is not often afforded to the crime scene officer or fingerprint technician and the investigating police are very much relied upon, along with the victim, for the provision of information that may assist in the examination.

At serious crime scenes a more extensive examination will generally be undertaken. This may include walls, bench tops or other large surfaces or items that may have been touched by the offender(s) during the commission of the offense. The use of some of the techniques available for the examination of these surfaces or items may in itself be destructive; therefore, the extent of their use must be considered with respect to the seriousness of the crime being investigated.

Any visible friction ridge impressions should be recorded photographically. Generally, it is not necessary to undertake any other means of improving the visualization of such impressions but a number of high-intensity forensic light sources are commercially available and may be taken to the crime scene to assist in examining the scene for fingerprints in the initial stages. These light sources are not generally carried by the examining officer when attending routine crime scenes, but are regularly used at serious crime scenes. When using them, it may be necessary to darken the scene by covering the windows, etc., or conducting the examination at night. An examination of a crime scene using a forensic light source may reveal fingermarks where the presence of some medium on the friction ridges responds to a particular wavelength in the light spectrum, enabling the mark to be visualized.

Routine examinations of latent fingermarks may be initiated with the simple use of a light, such as a torch, played over a surface, followed by the application of an adhesive development powder to those surfaces, considered suitable for this process. Generally, adhesive powders are only appropriate for the examination of smooth, nonporous surfaces, such as glass and painted wood or metal. The selection of powder, white, gray or black, depends upon the surface color, keeping in mind that the objective is to improve overall contrast of any impression located.

The powders are generally applied to the surfaces being examined with a light brushing technique. The brushes vary from soft squirrel hair to glass fiber. Magnetic powders are also available and are applied with a magnetic 'wand'. Once the latent friction ridge skin impression is visualized or developed, as is the term often used, it will be either photographed or lifted, using adhesive tape. In either case, the impression should be photographed *in situ* to corroborate its location, should this be necessary at a later time. A small label bearing a sizing graph and a unique identifier should be affixed adjacent to the impression for later printing of photographs at the correct and desired ratio. Should it be decided to lift the impression, this is then placed on to a suitable backing card, either black or white, depending on powder color, or on to a clear plastic sheet.

Many objects, particularly some types of plastics, and porous surfaces such as paper, are better examined using laboratory-based processes. Any such items located at the scene should be collected and sent to the laboratory for examination. For permanent fixtures at the scene not suitable for powder development, some laboratory-based examination processes can, with appropriate preparation, be used. Some of these techniques are addressed in the next section. It must be remembered that, in using these processes at the crime scene, issues relating to health and safety must be considered and jurisdictional regulations must be adhered to. In deciding which processes to use, sequential examination procedures must be observed when attempting to visualize friction ridge skin impressions, whether at the scene or in the laboratory.

Before departing from the crime scene, it will be necessary for the crime scene officer, with the assistance of investigating police, to arrange for inked finger and palm prints to be obtained from all persons who may have had legitimate access to the scene or items to be examined. These inked impressions are generally referred to as elimination prints and should not be confused, as is often the case, with inked impressions that may be obtained from suspects for the purpose of eliminating them from police inquiries. Such impressions are, in reality, obtained for the purpose of implicating rather than eliminating.

Laboratory Examinations

The most common methods of examining porous surfaces, such as paper, use chemicals such as DFO (1,8-diaza-9-fluorenone) and ninhydrin. These chemicals react with amino acid groups that are present in perspiration. If DFO and ninhydrin are both to be used, it is necessary to use DFO first for best results. Friction ridge skin impressions are generally difficult to visualize after processing with DFO; however, using a suitable high-intensity light source, in combination with appropriate filters, developed impressions will be photoluminescent, enabling photographic recording. Processing with ninhydrin will generally enable any impressions present to be

visualized, in ambient light, as a purple image. Once again, these can be recorded photographically.

If it is known that the paper items to be examined have been subjected to wetting then neither DFO nor ninhydrin is suitable. The reason for this is that amino acids deposited into the porous surface, being water-soluble, will be diffused. In such cases, however, a physical developer process can be used. This process relies on nonsoluble components of the friction ridge skin deposit, which may have been collected through touching other areas of the body or foreign surface, still being present on the surface touched. This process is more labor-intensive than methods using DFO or ninhydrin, but is quite effective. The impressions which may develop can be seen as very dark gray images and are easily visualized on lighter colored papers.

Items such as plastic, foil, firearms and knives can be processed using cyanoacrylate esters, commonly known as superglue. The items to be examined are placed into a closed chamber and a small amount of superglue is added. Heating the superglue helps it to fume. The glue will polymerize on to any fingerprint deposit which may be present on a given surface, appearing white and generally stable. Should the surface being examined be light in color, it will be necessary to enhance the developed impression with the use of dyes, stains or even powder. Once stained or dyed, superglue-developed impressions, like DFO, can then be visualized with the use of a high-intensity light source and appropriate filter, and then recorded photographically.

Evidence Evaluation

At the completion of the examination undertaken by the scenes of crime officer or fingerprint technician, the resulting adhesive fingerprint lifts or photographs will be forwarded to a fingerprint specialist. This specialist will examine the impressions submitted and conduct comparisons with known persons or search existing records in an effort to establish a match.

On receiving the evidence, the fingerprint specialist will undertake an evaluation of the quality of the impressions submitted. While a cursory evaluation is made at the crime scene, it is generally more definitive in an office environment with the appropriate lighting and equipment. Should it be determined that there is a significant lack of quality and quantity in the friction ridge skin impression, it may well be deemed, in the opinion of the specialist, to be of no value for any meaningful comparison. It is extremely important, however, that the evidence be retained with the case file for record and presentation, should full disclosure of evidence be required.

The remainder of the impressions would then be subject to comparison with known 'inked' finger and palm impressions. These inked impressions are obtained by placing the desired area of friction ridge skin on to black printer's ink and then placing the inked skin on to suitable white paper. Usually, this would be a blank fingerprint form.

Identification Methodology

The premises

Fingerprint identification relies on the premise that detail contained in the friction ridges is unique and unchanging. The premises or fundamental principles of fingerprint identification are:

- Friction ridges develop on the fetus in their definitive form before birth.
- Friction ridges are persistent throughout life except for permanent scarring.
- Friction ridge patterns and the detail in small areas of friction ridges are unique and never repeated.
- Overall friction ridge patterns vary within limits which allow for classification.

The fourth premise is not strictly relevant to individualization, as classification only assists in narrowing the search by placing the fingerprints into organized groups.

The analysis

In analysing the friction ridge skin impression, the fingerprint specialist will consider all levels of detail available. This may simply be through consideration of pattern type, should this be obvious in the impression located at the crime scene. An example of this would be where an arch pattern, being the crime scene impression, is compared with known impressions, none of which can be seen to be arches, in which case the comparison may well be concluded at this time. This pattern evaluation may be referred to as first level detail and can often be undertaken with little use of a magnifying glass.

Should pattern types be consistent, or not be entirely obvious, which may generally be the case, it will then be necessary to commence examination of the ridge endings and bifurcations in an effort to determine whether the impressions were made by one and the same person. These ridge endings and bifurcations, or 'traditional' characteristics may be referred to as second level detail. Often this is sufficient, and historically, this is the only detail contained in the impressions that would generally be used.

Establishing identity with 'traditional' characteristics

In determining whether two friction ridge skin impressions have been made by one and the same person, it will be necessary to locate sufficient ridge endings and bifurcations in true relative position and coincidental sequence. To achieve this, an examination of one of the impressions, generally that recovered from the crime scene, will be initiated. One, or perhaps two, characteristics may be observed and compared with characteristics appearing in the second impression, which would usually be the inked impression. Once such characteristics are located to the satisfaction of the examiner, a further examination of the first impression will be made in an effort to locate any additional characteristic. Again, the second impression will be examined to locate that characteristics. Having done this, the examiner will insure the characteristics observed are in the correct sequence, which means that each of the characteristics must be separated by the same number of intervening ridges. In addition, their position relative to each other must be considered.

An example of the process can be seen in **Fig. 4**. In the two impressions shown in the figure, it can be seen that the marked characteristics are in relative position with one another. It can also be seen that the number of intervening ridges between points, for example 1 and 2, are the same in each impression – five intervening ridges in this case. Likewise, there are three intervening ridges between points 2 and 3. This process is continued throughout an examination until the examiner is either satisfied that an identification has been established, or that the impressions are not identical.

As an exercise, continue counting the intervening ridges between the marked points. The following is a selection against which you may compare your own findings:

points 4 and 6 = five intervening ridges;
points 6 and 9 = nil intervening ridges (trace the ridge from point 6 and you will arrive at point 9);
points 10 and 8 = eight intervening ridges.

Numeric standards

As stated, fingerprint specialists have historically relied on the ridge endings and bifurcations for individualizing friction ridge skin impressions. Used in the manner discussed, this has resulted in the creation of numeric standards; that is to say, standards were set establishing a predetermined number of characteristics to be present, in relative position and sequence, before a fingerprint identification could be positively established for court presentation. These predetermined numeric standards differed throughout the world and have shown variations from 7 to 16. Each country had their own reason for their preferred standard, although many were based on statistical models that demonstrated the probability of no two people having the same fingerprint.

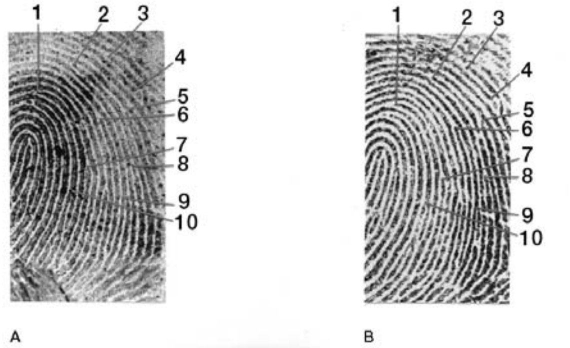


Figure 4 Friction ridge skin impressions: (**A**) crime scene; (**B**) inked. See text for discussion.

In 1973, the International Association for Identification (IAI) met for their annual conference in Jackson, Wyoming, USA. A statement declaring

> no valid basis exists at this time for requiring that a predetermined minimum of friction ridge characteristics must be present in two impressions in order to establish positive identification,

was subsequently adopted by all North American fingerprint identification examiners. This began a philosophy whereby an opinion of identification was not based on the number of characteristics present in two impressions.

New South Wales Police, in Australia, followed this direction in the early 1980s and, in doing so, broke away from the national standard, which was 12. The Australian Federal Police and the Northern Territory Police (Australia) also soon adopted a nonnumeric philosophy. In 1998, it was agreed that the national standard in Australia be accepted as nonnumeric; however, state jurisdictions are at liberty to maintain 'office policy' regarding minimum standards.

The United Kingdom maintains a numeric standard of 16 friction ridge skin characteristics before an identification can be presented at court, in routine circumstances; this is considered by many to be extremely conservative. This standard was challenged in a report prepared in 1989, which, however, was only released to a meeting in Ne'urim, Israel, in June 1995. The report was soon published in *Fingerprint Whorld*. The following resolution was agreed upon and unanimously approved by members at that meeting:

> No scientific basis exists for requiring that a predetermined minimum number of friction ridge features must

be present in two impressions in order to establish positive identification

Note the significant difference in the resolution, *'no scientific basis exists ...'*, compared with the IAI declaration, *'no valid basis exists ...'*.

The detail contained on and along the friction ridges, such as the sweat pores and ridge edges, are as unique to the individual as are the combination of ridge endings and bifurcations and may be used to support a fingerprint identification. This detail may be referred to as third level detail with its uniqueness soundly based in science. Third level detail is being increasingly accepted and used by fingerprint specialists rather than simply relying on a predetermined number of bifurcations and ridge endings. The scientific application of all the unique detail contained in the friction ridge skin is now being encompassed in a form of study being referred to more commonly as forensic ridgeology.

Ridgeology

Structure of friction ridge skin

To assist in understanding ridgeology, a description of the structure of friction ridge skin may be a good point at which to begin. This skin comprises two primary layers: the inner, or dermal, layer; and the outer, or epidermal, layer. Within the epidermal layer are five sublayers, the innermost layer being the basal. This layer connects to papillae pegs, which are visualized on the outer surface of the dermis. The basal layer generates cells that migrate to the surface of the skin. This migration, among other functions, enables the persistence of the detail visible in the friction ridges. From within the dermal layer, sweat ducts, from the eccrine gland, twist their way to the tops of the friction ridges, where small sweat pores may be seen.

The friction ridges themselves are constructed of ridge units, which may vary in size, shape and alignment, and 'are subjected to differential growth factors, while fusing into rows and growing'. The ridges are three-dimensional, creating a uniqueness in the formation of the friction ridges, even in a very small area. Random forces, which result in differential growth, also affect the location of the sweat pore openings within a ridge unit. The detail associated with the sweat pores and the minute detail located on the ridge edges is significantly smaller than the traditional characteristics. It is however the minute detail and differential growth, which determines the position of the ridge endings and bifurcations, and overall ridge shape.

Poroscopy

Poroscopy is the method of establishing identity by a comparison of the sweat pores along the friction ridges; it was extensively studied by the French criminologist, Edmond Locard, who found that sweat pores varied in size from 88 to 220 μm. Locard demonstrated the value of poroscopy in the criminal trial of Boudet and Simonin in 1912, in which he marked up some 901 separate sweat pores and more than 2000 in a palm print recovered from the crime scene. Other such cases involving much lower numbers of sweat pores have been documented in more recent times.

Edgeoscopy

The term edgeoscopy surfaced in an article which appeared in *Finger Print and Identification Magazine* in 1962. The author suggested the use of ridge edges in conjunction with other friction ridge detail, which may be present, and assigned names to seven common edge characteristics, which included convex, concave, peak, pocket and angle.

Additional friction ridge skin detail

Further detail that may be utilized by the fingerprint examiner includes scarring, usually of a permanent nature, and flexion creases. In examining the palm area of the hands, the predominant flexion creases can be seen to be the point where the skin folds when you start to make a fist of your hand. The persistence of these creases was demonstrated by Sir William Herschel who compared two separate impressions of his own left hand taken 30 years apart.

Two murder trials in which flexion crease identifications were made have been conducted in North America.

Permanent scarring is created by injuries which may be inflicted upon the inner extremities of the epidermis, effectively damaging the dermal papillae. Once such a scar is created, it too becomes a permanent feature of the particular area of friction ridge skin. This permanence allows for such a scar to be used in the identification process when it appears in the two impressions being compared.

Fingerprints and Genetics

The friction ridge detail appearing on the palmar and plantar surfaces of identical, or monozygotic, twins will be as different and varied as will be encountered in each of us. It must be remembered that it is the detail located within and along the ridges that enables us to individualize areas of friction ridge skin. The pattern the ridges form, however, may very well be

influenced by genetics. A number of publications have resulted from the study of the development of friction ridge skin. These studies were lengthy, so no attempt to disseminate their contents will be made in this discussion; however, a brief and very simplified account of their outcome is provided.

In the development of a human fetus, volar pads form and become evident around the sixth week. The placement of the volar pads conform to a morphological plan, and the placement and shape of these pads, which are raised areas on the hand, may influence the direction of the friction ridge flow, which in turn creates the particular pattern types that may be seen. Examples of this, it is believed, would be high cent-red pads, creating whorls, while intermediate pads with a trend are believed to form loops. While genetics may influence the pattern formation and general flow of the ridges, the location of ridge endings, bifurcations and other detail contained within each ridge is the result of differential growth at the developmental stage.

Conclusion

Fingerprints have been used for personal identification for many years, but not as much as they have throughout the twentieth century. It has been scientifically established that no two individuals have the same friction ridge skin detail, nor will one small area of this skin be duplicated on the same individual. This fact will not alter, nor therefore, will the importance of the fingerprint science and the role it plays in the criminal investigation process.

See also: **Fingerprints (Dactyloscopy)**: Visualization; Sequential Treatment and Enhancement; Identification and Classification; Standards of Proof; Automated Methods, including Criminal Record Administration.

Further Reading

Ashbaugh D (1999) *Quantitative – Qualitative Friction Ridge Analysis. An Introduction to Basic and Advanced Ridgeology*. Boca Raton, FL: CRC Press.

Barclay F (1997) Friction ridge identification based on pore structure: a case study. *Fingerprint Whorld*, **23**: pp. 7–11.

Clegg D (1994) Poroscopy in practice. *Journal of Forensic Identification* **44**:15–20.

Cummins H and Midlo C (1943) *Finger Prints, Palms and Soles. An Introduction to Dermatoglyphics*. Boston, MA: Research Publishing.

Evett IW and Williams RL (1995) A review of the sixteen points fingerprint standard in England and Wales. *Fingerprint Whorld* **21**:125–143.

Margot P and Lennard C (1994) *Fingerprint Detection Techniques*, 6th edn. Lausanne: Institut de Police Scientifique et de Criminologie.

Nelipa P (1997) *A study of crime scene management strategies, evidence collection and evidence examination techniques in the USA, UK, Republic of Ireland, Germany, Israel and Russia*. Report to Winston Churchill Memorial Trust: Australian Federal Police.

Police Scientific Development Branch (1998) *Manual of Fingerprint Development Techniques*, 2nd edn. London: Home Office.

Major Incident Scene Management

J Horswell, Australian Federal Police, Canberra, Australia

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0458

Background

There are three critical operational stages of a criminal investigation. These are:

- control and coordination of the criminal investigation;
- the criminal investigation;
- the forensic investigation.

Normally the coordination of a major criminal investigation is delegated to a senior investigating officer; likewise a complex forensic investigation must be coordinated by a senior forensic investigator. To manage major incident scenes and multisited crime scenes the crime scene coordinator/manager must understand forensic science and criminal investigation. The crime scene manager must surround him- or herself with competent, skilled and qualified forensic investigators, not prima donnas.

This article only deals with the management of major incident scenes from a forensic investigator's point of view.

Scene Control and Coordination

Without proper control and coordination, information may not reach the crime scene investigator. This may lead to his or her efforts being aimless, and the leads uncovered may never be passed on to to investigators for follow-up action. This is most important when the crime scene is large and there are several crime scene investigators present processing the scene,

or where there are secondary scenes away from the primary scene. There must be a flow of information back and forth between the senior investigator and crime scene investigator. This is one of the functions of the crime scene manager.

Approach to Crime Scene Investigation

The examination of a crime scene and subsequent collection of potential physical evidence requires special skills, knowledge and aptitude. The manner in which a crime scene examination is conducted may be a critical factor in determining the success of an investigation. The proper examination of a crime scene requires a disciplined approach and systematic application of the various observation, recording and collection techniques, as well as an in depth knowledge of forensic science.

Examining a crime scene is often a demanding task, and in many instances requires physical and mental stamina as well as team leader/member skills.

Forensic science has become a powerful aid to criminal investigations, with courts placing much emphasis on the results. Accordingly, the manner in which evidence is collected, the observations made and the results of tests and comparisons conducted are vigorously examined by the courts.

A systematic approach to crime scene investigation will ensure:

- good coordination between investigation and crime scene examination teams;
- an efficient, effective and thorough examination;
- less fatigue;
- orderly recording and collection of potential evidence;
- effective observations and deductions.

Initial Assessment

Before attending the crime scene it is important to obtain the best possible assessment of the circumstances relating to the crime. It is also important to receive a briefing from the senior investigating officer who has been appointed to conduct the investigation. From a forensic viewpoint, a crime scene coordinator should be appointed. This person will remain the focal point of contact between all the players who will subsequently become involved in the forensic investigation. This person should be a senior crime scene investigator who will be responsible for chairing all subsequent meetings with investigating police and for the coordination of all aspects of the forensic investigation. This includes the allocation of human resources to multiple scenes.

Homicide will be discussed here as the model crime scene as this is the most serious crime against the person.

Forensic investigators are in the same position as investigating officers. They need answers to the same questions: Who? What? When? How? Where? Why? Some of these questions can be answered at the homicide scene.

Who?

- Who is the deceased?
- Who reported finding the deceased?
- Who saw or heard anything?

What?

- What happened?
- What crime, if any, has been committed?
- What were the actions of the deceased?
- What were the actions of others involved?
- What are the witnesses saying?
- What injuries, marks, clothing and personal property are on the deceased?
- What is the estimated time of death?
- What possible weapons were used?
- What cause of death information can be gleaned from the deceased?
- What was the manner of death?

When?

- When was the deceased discovered?
- When was the deceased last seen alive?
- When were the police notified?

How?

- How did the deceased get to the death scene?
- How long has passed between injury and death?
- How did the deceased sustain the injuries?

Where?

- Where was the body discovered?
- Where did the death occur?
- Where was the deceased last seen?
- Where did the injury/ies occur?
- Where were the witnesses during the incident?

Why?

- Why was the deceased at this location?
- Why was a crime committed?
- Why was this type of weapon used?
- Why did this death occur at this time?
- Why was the deceased found at this time?

The more detail that can be obtained about what happened, the easier it is to determine what resources are required in examining the scene and

the significance which is to be placed on certain aspects of the evidence.

Scene Security

An initial assessment of the crime scene will be made by the first officers attending the scene. The scene will be secured by them to an extent based on the information available at the time. The crime scene manager, who will normally be a senior member of the crime scene staff, must attend the scene at the earliest possible opportunity to take charge of the management of the crime scene. He or she will usually be accompanied by a crime scene investigator or a team of crime scene investigators who will undertake the crime scene examination. The size of the crime scene/s will dictate the amount of resources allocated to the particular incident. It is imperative that the crime scene manager has the authority to allocate the amount of resources required.

Once the crime scene is handed over to the crime scene manager, a reassessment of the scene security should be made to ensure that it is adequate. There should be a formal protocol used for the handing over of a crime scene. This ensures control and the maintenance of the scene's chain of custody.

It is an essential element of any prosecution where forensic evidence is involved to prove the security of the scene and that it was maintained throughout the subsequent examination/investigation. Therefore the objectives of securing the crime scene are:

- To prevent evidence being destroyed or contaminated.
- To insure security of information; generally information is only released to the media by a media liaison officer or the senior investigating officer.
- To insure chain of custody of the scene is maintained as is necessary with any item of potential evidence.
- To remove from the scene all unnecessary persons including police officers and the media. It must be remembered that the more people present, the greater the potential for contamination and destruction of evidence. Large numbers of persons present will also inhibit the proper processing of a crime scene.
- To insure that all evidence has been recorded and recovered. This may include securing the scene until the results of the postmortem or scientific analysis are to hand.

There are a variety of methods for securing the crime scene, for example:

- posting guards;
- rope or printed tape cordons;
- the strategic placing of vehicles;
- the use of markers, flags, signs;
- locking rooms or areas within buildings or using the external walls of a building as the barrier;
- establishing safe walk areas (common approach path) with tape or purpose-built raised stepping plates.

Occupational Health and Safety

The well-being of the crime scene investigator/s is the primary responsibility of the crime scene manager. He or she must be aware of fatigue and well-being of the investigators. Appropriate protective clothing and equipment should be made available. Breaks should be organized for the forensic investigators and refreshments should be on hand during those breaks. Scene guards should also be part of the crime scene operation, regardless of the area they originate from. There should be an area designated where food and drink can be taken, equipment can be stored and rubbish can be accumulated.

All personnel on site should be briefed regarding:

- safety hazards;
- smoking and eating;
- the location of critical areas;
- the use of telephones and toilets.

Systematic Collection of Potential Evidence

After the general survey of the crime scene, the sequence in which evidence is to be collected and areas searched should be apparent. Priority should be given to:

- any items that are in danger of being destroyed by wind, rain, vehicles, animals, tides and human movement;
- the collection of any evidence that will enable access to a deceased or any critical area of the crime scene, such as along entry and exit paths;
- those critical areas of the crime scene that may render the most evidence or, once processed, enable the removal of a body, or the remainder of the examination to begin;
- any area that may give a quick indication as to the identity of the offender/s;
- any areas that, when processed, will permit the release of scene guards and other resources;
- a general examination of the remainder of the crime scene for potential evidence.

Systematic and Sequential Approach to the Search and Recovery of Potential Evidence

In establishing the manner and sequence of collecting potential evidence, consideration must be given both to the possible destruction of evidence and to the approach which will yield the best result in terms of useful information. Consultation with other specialists, such as forensic scientists and forensic pathologists, as to the sequence and method of collection may be necessary to ensure the best result; however, this may not always be possible at the scene.

Some examples of collection procedures are as follows:

- Macroscopic evidence should be collected from an area before it is powdered for fingerprints.
- Bloodstains and forensic evidence should be collected from an area before searching for fingerprints.
- Sweepings from the floor around a safe need to be collected before the magna brush is used.
- Polished floors need to be examined first with oblique lighting to locate latent shoemarks/footprints.
- Visible fibers, hairs and other trace material should be collected from an area before applying general collection techniques, such as tapelifts, sweeping and vacuuming.
- Tapelift areas of interest before removing deceased persons, collecting sheets and blankets.

In searching critical areas, a search conducted in a clockwise or anticlockwise direction from a fixed point, or conducting a line strip search, makes for a systematic examination. A systematic approach reduces fatigue and ensures a more comprehensive search by minimizing the chance of missing potentially valuable evidentiary material.

Larger objects should be examined before smaller objects and all items should be packaged and labeled at the time of collection.

Examination Records

In order to conduct a thorough, systematic crime scene investigation a proforma should be developed for each activity. These will not be enlarged on here as each jurisdiction will have its own subtle differences, but a list of subject headings for each category of examination is given below. These should be preprepared checksheets that will provide the examiner with comprehensive notes taken during the examination; these proforma records should be available for:

- crime scene log – activities undertaken at the scene, including movements in and out of the scene;
- formal handover of the crime scene;
- list of environmental conditions at the crime scene;
- list of activities and observations at the crime scene;
- exhibit list;
- rough sketch of the crime scene;
- photographs taken at the scene;
- list of specialists attending and times they were involved in the examination.

Ongoing Case Management

Once the scene work is completed, the emphasis changes to the coordination of further examinations and the communication and flow of information of the results from forensic examinations to investigators and from investigators to forensic examiners. If it is not practical for the crime scene coordinator to chair further case management meetings, another senior crime scene investigator may be nominated to maintain that contact and coordination of the ongoing case management.

Summary

Management of major or minor crime is a matter of seizing control of the crime scene and then the coordination of resource management, along with a systematic approach to processing the scene.

Major crime scenes vary in size and complexity. Some may require many crime scene investigators; others which are uncomplicated may require only one or two crime scene investigators.

Overall scene management and the maintenance of a two-way communication flow are the two essential ingredients to successful scene management. Regular case management meetings must be held to keep all stakeholders abreast of the latest available information. These must be recorded in the case notes.

See also: **Crime-scene Investigation and Examination**: Recording; Recovery of Human Remains. **Health and Safety**: Including Risk Assessment. **History**: Crime-scene Sciences.

Further Reading

Fisher Barry AJ (1993) *Techniques of Crime Scene Investigation*, 5th edn. Springfield, IL: CRC Press.

Goddard KW (1977) *Crime Scene Investigation*. Reston, VA: Reston.

Horswell J (1993) Crime scene investigation. In: Freckelton and Selby (eds) *Expert Evidence*. North Ryde, NSW: Law Book Company.

Kirk P (1974) *Crime Investigation*, 2nd edn. Chichester: Wiley.
Svensson A and Wendel O (1974) *Techniques of Crime Scene Investigation*, 2nd edn. New York: Elsevier.

Packaging

J Horswell, Australian Federal Police, Canberra, Australia

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0452

Background

The ideal method of collecting and subsequent packing of items for transport to the forensic science laboratory will vary considerably depending on the nature of the item concerned. Likewise the material from which a package is fabricated will also vary considerably.

Paper as a Packaging Medium

Generally, the use of paper in the form of bags of various sizes is recommended.

Paper bags are fabricated in various sizes and should be readily available. If using envelopes, A4 white bond paper should be employed in the first instance for small items such as hairs, fibers, glass or paint. Place the item onto previously folded paper and then place this in an envelope. This will prevent the loss of items from envelope corners, and the use of folded paper will simplify the subsequent examination under a macroscope.

The placement of clothing and biological material in paper allows the item to breath; placement in a plastic bag may result in bacterial action and encourage the growth of mold. All items should be air dried prior to packaging in paper.

Extensively blood-stained or semen-stained items recovered from crime scenes should be first placed into paper and then protected by a plastic bag; however, the item must be removed from packaging on arrival at the forensic science laboratory and air dried.

Items wet with volatile substances should be placed in nylon bags or new, clean, paint cans. Normal polyethylene bags are not suitable for the retention of volatiles.

Infested Material

In some instances material recovered from crime scenes or mortuaries for return to the forensic science laboratory, or for long-term storage as property, may be infested with pests, such as fleas, lice, maggots or coffin beetles. Care must be taken when examining this material by wearing protective clothing and gloves. If possible, always use a large open search bench.

If insect infestation is present within the item there are at least two methods available for killing them:

- Place the material and container in a large plastic bag and seal it. Place the bag into a deep freeze until the insects are dead.
- Add a few drops of ethyl formate to the plastic bag containing the item and its container. Seal the bag and leave for approximately 1 h or until the insects are dead.

Where blood or semen stains are present, samples from stains must be collected prior to freezing. Consideration should also be given to entomological aspects of the case: both live and dead insect specimens may be required for examination.

Collection of Items

Quantity It is better to collect excess material than to have an examination fail because there is insufficient material for analysis. Where difficulty may be encountered in collecting minute traces of substances, specialist collection techniques should be employed. If, however, traces of evidence are on small items and there is a chance of loss, the trace subsample should be recovered and placed into a separate package. If the trace is a stain then the stain should remain on the item for assessment in the forensic science laboratory. This is particularly relevant where the item as a whole is vitally relevant; for example, a blood-stained knife.

Controls In many cases involving stained material, for example fire debris, it is necessary to submit unstained material for analysis to determine if the material itself interferes with the analytical procedures.

Reference material In any comparison of the constituents of two substances, sufficient reference sample material should be provided. For example, if dust on clothing is suspected of being ballast from a particular safe, a sufficient amount should be collected, packaged and submitted together with the items of clothing in order that a satisfactory comparison may be performed. The reference sample should be repre-

sentative of the source from which the test sample originated. For example, it is useless attempting to compare a body hair found at the scene with a head hair from the suspect. Similarly, the comparison of soil from a shoe with soil from the scene may be unsuccessful if the sources of the two samples are separated by only a few meters.

Labeling

The purpose of a label is twofold:

- to identify the nature and source of the item;
- to establish a chain of custody.

Ideally, a label should have the following information recorded on it:

- nature of contents;
- source (where found or from who recovered);
- date and time;
- signature and printed surname of collector;
- sequential number;
- unique case identifying number.

Additionally, room should be available on the label to record the movement of the item.

The label should be completed at the time of collection or receipt of the item.

If an item is collected from a person, use that person's name; for example, 'Trousers from John Smith'. Do not mark the item with the word 'Suspect', as this may lead to the exhibit being excluded from a trial. Some courts are of the view that to mark items in this way is 'unnecessary' and 'objectionable' because whenever such an item is mentioned in evidence during the trial the jury are being told that the accused has been a suspect. This perhaps gives the impression that he or she was a suspect early in the investigation, which need not necessarily be the case. The jury may also wrongly hold the view that there may be more to know about the accused, which could be prejudicial. Obviously the word 'offender' should never be used at this is a presumption of guilt.

The sequential number used should relate to the collectors item list and could be JH1, JH2, JH3, etc. When making a subsequent examination of the items collected, any material removed should be given a number that relates to the original item. For example a pair of trousers is marked JH1 and hair is recovered from the trousers. This item should be marked JH1.1. In this way each subsample can be traced easily back to the original source item.

Where the item is something substantial, for example a knife or clothing, then the exhibit itself should be marked as well as the container. It may be appropriate to tie a label with a piece of string to the item. If this is done then there can be no doubt about later identifying the item in the witness box should it be separated from its container.

If using plastic pots or vials, ensure that there is a corresponding mark on both the lid and the container to avoid any mixing up of the containers. The sequential number and unique case number are normally used for this purpose.

Collection

The proper collection of items and trace material is essential in obtaining the greatest evidential value from an examination.

Special clothing should be worn during all scene and laboratory examinations. Scene suits, white cotton overalls or laboratory coats should always be worn as white cotton has the least evidential value as a fiber and is therefore suitable in preventing contamination of scenes or clothing with fibers from the examiner's clothing. There is also an occupational health and safety dimension to the use of appropriate clothing.

Collection case

Collection cases must be kept clean, with equipment stored in an orderly manner. The principal collection items and their uses are listed in **Table 1**.

Collection Techniques

A variety of techniques have been developed for the collection of trace material and other potential evidential material. Each technique is designed to prevent damage to, and contamination of, the material. The main collection techniques can be described as:

- handpicking;
- tape lifting;
- swabbing;
- sweeping;
- vacuuming.

Handpicking

Whenever examining the crime scene, garments, bodies or other articles, the initial emphasis should be directed toward the collection of gross and macroscopic items that can be recovered by hand or by the use of tweezers. Items large enough to see with the naked eye should be collected by handpicking. Material such as hairs, large paint and glass fragments and pieces of vegetation should be collected before the application of other collection techniques, such as tapelifting, sweeping and vacuuming.

Handpicking has the advantage of establishing the position of the material on the item and requires no further time in searching, whereas tapelifts, sweepings and vacuumings have to be further searched to

Table 1 Principal items of collection equipment and their uses

Item	Collection use
Scalpel	Paint smears, visible fibers, vegetation and dried blood
Probe	Paint, fibers, residues, oils, greases; manipulation of microscopic particles
Brush	Trace particles: paint, metals, vegetation, glass
Swab	Small particles, which will be caught in the coarse fibers of the swab
Paint brush	To sweep localized and constricted areas
Spatula	Soil samples, whole or partly congealed blood; mixing of casting compound
Tweezers (metal)	Trace material, such as fibers, hair and vegetation
Tweezers (plastic)	Items that may be damaged if metal tweezers are used; recovery of bullets and fragments during postmortem examinations and for use when recovering blood stains using small pieces of moistened white cotton. Employed where cleaning is not possible and a number are required. Each tweezer is inexpensive therefore can be destroyed after each use
Cotton	Dried blood linen stains once moistened with distilled water
Magnet	Particles of iron and steel, after covering the magnet with plastic

isolate small particulate matter of interest.

When collecting items by hand, consideration should be given to the potential for contamination by perspiration and, where applicable, gloves should be worn.

Various types of tweezers are available to cope with small particulate matter, and a moistened fine hair brush will recover paint particles.

It is essential that each item of collection equipment is cleaned between individual collections.

Swabbing

Dry swabs can be employed to collect minute particles. The fibrous nature of the swab end can be effectively used to collect particulate matter. The material can then be separated, in the laboratory, from the swab on to microscope slides for further microscopic examination.

Swabs moistened with distilled or injection water can be used to collect body fluids. Forensic biologists used to prefer body fluids to be collected using cotton threads or small pieces of cotton material. The reason for this is related to the dilution and dispersion of the sample. For early electrophoretic techniques localized concentration was important and therefore the thread-type swabbing method was preferred. Now with DNA polymerase chain reaction (PCR) amplification, the dispersion and dilution of body fluids found on a moistened swab no longer present a problem.

Tapelifting

Tapelifting is a reliable method of collecting trace microscopic material from a variety of surfaces: in particular, garments and motor vehicle seats. Transparent adhesive tape no more that 7.5 cm in length is applied to the surface of the object. At the completion of the application the tape is placed over a clean piece of glass or rigid plastic and then placed into a clean labeled plastic bag. Garments and other larger objects should be examined in segments; for example, the front and rear of a shirt as two discrete areas. The tape should only be used while the adhesive qualities remain.

Too much material should not be placed on one tape. The collection of material in this manner facilitates the examination of trace evidence using a macroscope and, in particular, assists in sorting material of interest from a myriad of other insignificant material. When using adhesive tape from a dispenser, the first 5 cm should be discarded to prevent contamination. The method of tapelifting is used more widely in the forensic science laboratory, although it does have its uses in the field.

Sweeping

This method is particularly useful in collecting material from a variety of areas, including inaccessible sites or those where there is a mass of material. Sweeping is also a useful collection technique for the examination of motor vehicles where large amounts of debris can be present on the floor surfaces, in boots or in engine compartments.

It is essential that the brush is clean and that separate brushes are used whenever contamination or cross-transfer is a consideration; for example, examining a scene and a suspect's vehicle. New paint brushes approximately 25 mm wide with non-painted handles, along with new pans from dustpan and broom sets, should be used on each occasion where sweeping is employed.

Vacuuming

The collection of microscopic material, from garments, motor vehicles and other large objects, by vacuuming is another valuable means of collecting

trace material. However, the circumstances in which it should be employed need to be considered carefully, as the vacuumings collected are difficult to handle, involving the expenditure of a great deal of time in searching them in the laboratory. Furthermore, vacuuming can be too effective, in that it can lead to the collection of a great deal of 'ancient history'.

This method requires a specialized nozzle for the vacuum cleaner. Nozzles are custom made from stainless steel. Plastic demonstration traps have also been employed in the field.

Material is collected by suction on to a clean filter paper (stainless steel) or cotton material (plastic), which rests on a perforated plate located in the central compartment. Some situations may warrant the use of an ordinary vacuum cleaner with a clean bag for collecting trace material; however, this is a last resort as you will add material to the collection sample from traces which remain in the vacuum hose.

When using the stainless steel and plastic traps, care must be taken to ensure that the nozzle, trap and suction end of the traps are cleaned before use, or when vacuuming separate articles or particular localized regions of an object, vehicle or scene. The nozzle should be washed in warm soapy water, rinsed with clean water and dried. Bottle brushes are ideal for cleaning nozzle pipes. When in the field away from the ability to clean the nozzle it must be brushed clean between each use.

A blank/control vacuuming should be run before the actual sampling run, with a clean filter paper in place. This is then removed and bagged separately for later examination. Each sample run must also have a clean piece of filter paper, likewise a new bag as a control and new bag for the sample run if a trap nozzle is not available.

Once set up ready for a sample run, the nozzle is applied to the surface; for example, with a garment in a series of strokes. Each area of the garment will be a discrete searching area in its own right; for example, pockets, back and front of garment.

When not in use the nozzle should be cleaned and stored in a sealed plastic bag.

Preservation

Items must be preserved so that they remain, as far as possible, in their original state and may be produced in court in the condition in which they were found. In some cases it is not possible to retain the exhibit intact; for example, in analytical procedures the item may have to be altered or it may be totally consumed in the procedure.

The crime scene investigator should take all necessary steps to protect items collected from the following:

Loss Small items such as hairs, fibers and paint flakes may be lost from packages that are not properly sealed. Envelopes on their own are unsuitable for small samples as the particulate matter may be lost through corners of the envelope. Volatile liquids from fire scenes may evaporate if the containers are not airtight and impermeable.

Deterioration or damage Biological material such as wet blood or seminal stains may deteriorate rapidly. Valuable shoe impressions and blood stains in outdoor scenes must be protected and collected before wind and rain destroys them.

Contamination Items which are not properly packaged may become contaminated by the introduction of foreign matter into the packaging.

Tampering Items should be packaged securely and should not be left unattended. The crime scene investigator should guard against innocent tampering as well as that intended to destroy potential evidence; for example, a firearm left unattended with a fired cartridge case in the breech may arrive at the forensic science laboratory with several impressions on the firing pin if the firearm is not packaged appropriately and secured.

Sealing Containers

The sealing of containers is necessary to keep items from being lost, contaminated or tampered with. The container should be sealed with sealing tape and then with evidence tape. The evidence tape should be signed by the crime scene investigator/collector.

Chain of Custody

The chain of custody refers to the documentation of possession of items from their recovery collection through examinations to their tendering in court as potential items of evidence. This allows interested parties to trace who has had custody of the item at a given time, as well as being able to account for where the item has been while it has been in an individual's or organization's custody.

Proximal containers, and, if applicable, items should be labeled with a movement record of the container/item, and the case file and/or the exhibit movement log should also record the movement of the item.

Hazard Labeling

It will be necessary to mark containers with appropriate hazard labels. Those that contain items which are stained with body fluids should be marked with a biological hazard label, and those that contain items that have been treated with chemicals to enhance fingerprints should be marked with a chemical hazard label. This should go some way in encouraging court staff to resist the temptation to open all packages and dispense with containers before the item is tendered as an exhibit in court.

Summary

This article has discussed packaging material and how to deal with infested material, covering the techniques employed and the sequence of collecting items.

Chain of custody has been covered, as has the use of appropriate labeling and sealing. Appendix 1 outlines the collection of specific items normally encountered in crime scene investigation, pointing out the most appropriate packaging and collection technique and the significance of each specific item.

See also: **Crime-scene Investigation and Examination**: Recording; Collection and Chain of Evidence; Recovery of Human Remains; Preservation; Contamination; Major Incident Scene Management.

Further Reading

Goddard KW (1977) *Crime Scene Investigation*. Reston, VA: Reston.

Fisher BAJ (1993) *Techniques of Crime Scene Investigation*, 5th edn. Boca Raton, FL: CRC Press.

Forensic Science Service (1999) *The Scenes of Crime Handbook*. Birmingham, UK: Forensic Science Service.

Horswell J (1993) Crime scene investigation. In: Freckelton I and Selby H (eds) *Expert Evidence*, pp. 8-7301–87611. North Ryde, NSW: Law Book Co.

Kirk P (1974) *Crime Investigation*, 2nd edn. Chichester: Wiley.

Svensson A and Wendel O (1974) *Techniques of Crime Scene Investigation*, 2nd edn. New York: Elsevier.

Appendix 1 Collection and Packaging

The following is offered as a guide to the collection and packaging of commonly encountered items of physical evidence.

Ammunition

This includes projectiles, live and expended cartridges, shot and wads.

1. Wrap with tissue paper, or in the case of distorted projectile fragments in a small plastic bag, and place each item in a separate plastic or cardboard container. Never mark ammunition – label the proximal container instead.
2. Wash projectiles and air dry if removed during autopsy.

Packaging: Item placed in a plastic bag then inserted into a rigid plastic or cardboard container. Do not use cotton wool or tissue paper for fragmented and distorted projectiles. Significance: Determine type of ammunition and its origin.

Bite marks

On skin

1. Photograph with colour and black and white ultraviolet film at 90° using a scale rule and gray card.
2. Wipe the area with a piece of cotton, moistened with distilled water, air dry, then place in a plastic tube.
3. Cast mark, if possible; place cast in small rigid plastic container.

Significance In consultation with an odontologist, comparison of bite mark with any suspect's teeth and possible DNA analysis.

On perishable items

1. Photograph.
2. Cast mark.

Packaging Place cast into a rigid plastic container.

Significance In consultation with an odontologist, comparison of bite mark with any suspect's teeth.

Suspect

1. Photograph teeth (five positions).
2. Obtain saliva using sterile cotton gauze, air dry and package.
3. Cast teeth of suspect; casts usually remain with the odontologist.

Packaging Place (2) in small rigid plastic container. Place (3) in small cardboard box.

Significance (1) and (3) are for use by the consulting odontologist for comparison work with bite marks and suspect's teeth; (2) for possible DNA analysis.

Blood

On absorbent material

1. Cut material, air dry and package separately.
2. Cut out a control sample.

Packaging Large blood-stained items should be

packaged in paper; small samples that have been dried should be placed in plastic phials and labeled.

Significance Blood grouping and/or DNA. Comparison of results with reference samples.

On nonabsorbent material

Wet For large volumes spoon or suck up with a disposable pipette liquid blood and package. For small volumes rub a small piece of cotton through the stain.

Dry For large volumes using a scalpel scrape into a small rigid plastic container. For small volumes rub a piece of previously moistened (distilled water) cotton through the stain and transfer the stain from the substrate to the cotton.

In both instances Controls from unstained area by moistening a piece of cotton and rubbing unstained area.

Packaging Plastic or glass phials.

Significance Blood grouping and/or DNA. Comparison of results with reference samples.

Whole blood

1. Obtained by doctor or nursing staff. Three samples: one straight sample and one with anticoagulant.
2. One with preservative.

Packaging For (1) glass or plastic phials. Anticoagulant: EDTA seeded pink top. For (2), placed in a preservative: sodium oxalate seeded brown top plastic phial.

Significance For (1), reference samples for blood grouping and/or DNA. Comparison of results with crime scene stains.

Bloodstain pattern interpretation

1. Use a large format camera, 6×6 cm minimum.
2. Photograph scene with black and white and color film.
3. Take overview photographs of stains at 90° from each subject.
4. Take close up views including measuring tape.
5. Take blood sample/s.

Packaging As above.

Significance Aid in the reconstruction of events and identify the donor of the blood stains.

Cigarette butts

1. Collect with plastic tweezers, air dry and package.
2. Examine for trace evidence: lipstick and fingerprints.

Packaging Plastic or glass phials. Package each butt separately.

Significance Identify cigarettes: more than one person smoking, DNA saliva residue, fingerprints and lipstick.

Clothing

1. Photograph, note and describe.
2. Remove clothing from suspects over clean white paper, air dry and package.
3. Remove any obvious trace material and package separately.

Packaging Paper bags. Place each item separately in a paper bag.

Significance Comparison with any trace material or blood stains with reference whole blood for proof of contact.

Documents

1. Wearing white cotton gloves, collect with tweezers and package.

Packaging Place in a prelabeled plastic document sheet or envelope containing a piece of cardboard.

Significance Determine origin, indented writing, fingerprints, obliterations or additions, or as material for comparison of handwriting in an attempt to identify the writer.

Fibers

1. Collect fibers with tweezers or submit whole item containing fibers.
2. Using the tapelifting collection technique, tape the area of interest using clear transparent adhesive tape and place tape, adhesive side down, on a clear plastic sheet.

Packaging Small plastic phial for loose fibers and plastic bags for plastic sheets.

Significance Identify possible source.

Fire debris

1. Collect debris from suspect area. Cut up large pieces of debris and package.
2. Collect control samples for different areas and package.

Packaging Clean new metal paint cans, nylon bags or PVDC (polyvinylidene chloride) bags.

Significance Determine the presence of and type of accelerant and distribution.

Firearms discharge residue

On hands

1. Photograph any visible evidence.
2. Collect from the web of the hand using adhesive SEM stubs.
3. Take control samples.

Packaging SEM firearm discharge residue collection kit.

Significance Determine if a firearm has been discharged by the person being tested.

On clothing

1. Photograph visible evidence and package.

Packaging Paper bag.

Significance Determine if a firearm has been discharged by the person who owns the clothes.

Bullet hole in clothing

1. Photograph with color and black and white (infrared) film.
2. Protect the bullet hole by pinning paper over it.
3. Place cardboard behind area to prevent bending. Do not bend clothing at the bullet hole if it can be avoided.

Packaging Paper bags.

Significance Determine distance from target.

Bullet hole in dead skin

1. Photograph with color and black and white (infrared) film.
2. Cut beyond the blackened area. surrounding the bullet hole and identify the '12 o'clock position' with a suture.

Packaging Glass jar and 10% formalin.

Significance Firing distance and angle.

Glass

At scene in general area

1. Photograph both sides of glass before removing it from the frame.
2. Collect fragile fragments first.
3. Wrap each separately to protect edges.
4. Place fragments in suitable container.
5. Collect all pieces if possible.

Packaging Plastic phial for small pieces for analysis; large cardboard boxes for larger pieces.

Significance Small pieces as control samples for refractive index measurement; identify source by mechanical fit and direction of breaking with larger pieces.

On clothing

1. Collect fragile fragments first.
2. Collect clothing and package separately.

Packaging Plastic phial and paper bags.

Significance Identify possible source by physical and/or chemical analysis.

Hairs

On moveable objects

1. Collect and protect the complete item.

On fixed objects

1. Using a pair of plastic tweezers collect fibers and package.

Control samples

Head Pluck 30–40 hairs from various areas.

Pubic Pluck 20–30 hairs.

Others Pluck 10–20 hairs from area of interest. Combed and plucked hairs are packaged separately.

Packaging Folded paper inserted into an envelope or plastic bag.

Significance Determine color, sex and race of person, areas of body, plucked or shed, human or animal.

Insects: flies

There are four stages in the life cycle of flies: eggs, maggots, pupae and adults (flies).

1. Collect 60–80 individuals from each position, on, under and approximately 90–150 cm from the decomposing body.
2. Collect from hidden areas, beneath leaves and floorboards.

Note Pupae may be found in the form of brown capsules under the body or in soil under the body.

Packaging Glass container. Place specimens in 70% V/V ethyl alcohol in distilled water.

Significance Estimation of time of death.

Maggots

1. Collect two samples, each containing 60–80 individuals.
2. Package one sample with some flesh.
3. Package second sample in alcohol or formalin.

Packaging Glass bottles.

Significance Estimation of time of death.

Paint

On tools or vehicles

1. If possible collect the item (tool or vehicle) containing the evidence.
2. Collect paint chips separately.
3. Care should be taken not to fragment paint chips.
4. Take reference samples of each color, ensuring they are scraped down to the base color.

Packaging Folded paper inserted into an envelope or plastic bag or rigid plastic container.

Significance Determine possible source: color, model, type of vehicle. Identify the vehicle when it comes to notice.

On clothing

1. Collect fragile evidence first.
2. Collect clothing and package individually.

Packaging Folded paper in an envelope or plastic bag and rigid plastic container.

Significance Determine possible source: color, model, type of vehicle. Identify the vehicle when it comes to notice.

Postmortem samples

Alcohol

1. Obtain clean arterial blood (10 ml).
2. Obtain urine (10 ml).
3. Obtain vitreous humor (10 ml).

Packaging Plastic tube containing a preservative (oxalate or fluoride).

Significance Indicates state of intoxication.

Blood for DNA

1. Obtain clean arterial blood (10 ml).

Packaging
Plain glass or plastic tube.

Significance Reference sample for comparison with scene and other unknown stains.

Blood for serology

1. Obtain clean arterial blood (10 ml).

Packaging Pink top plastic tube containing EDTA.

Significance Reference sample for comparison with scene and other unknown stains.

Drugs

1. Obtain clean arterial blood (10 ml).
2. Obtain urine (10 ml).

Packaging Plain glass or plastic tubes without anticoagulant or preservative.

Significance Indicate if under the influence of drugs or is a drug abuser.

Toxicology

1. Obtain clean arterial blood (10 ml).
2. Obtain a portion of liver (100 g).
3. Obtain all the stomach contents.
4. Obtain vitreous humor (10 ml).

Packaging Plastic containers.

Significance Poisons determination.

Diatoms

1. Obtain a portion of femur (15 cm).

Packaging Glass containing alcohol.

Significance To establish location of drowning: fresh water or salt water.

Safe insulation

From suspect's clothing

1. Collect clothing and package separately.

Packaging Paper bags.

Significance Comparison of any trace material found with a reference sample of safe insulation.

From safe

1. Collect reference sample of safe insulation from the safe.

Packaging Rigid plastic container.

Significance For comparison with any trace material found on suspect or his or her environment and therefore to link suspect with the scene.

Saliva

1. Collect saliva (ensure it is saliva and not mucus) on clean white gauze (wear gloves).
2. Air dry and package.

Packaging Rigid plastic container.

Significance DNA and determination of ABO group substances.

Seminal stains

1. Photograph stains and their distribution.
2. Collect items exhibiting stains (wear gloves).
3. Air dry and package.

Packaging Paper bags.

Significance Identification of donor by DNA and/or blood grouping.

Soil

1. Collect sample from suspect vehicle or footwear.
2. Collect clothing from suspect.
3. Collect several separate samples from scene and surrounding area as reference samples (50 g).

Packaging Rigid plastic containers for the soil and paper bags for the clothing.

Significance Geographical origin of samples, possible link between suspect and scene.

Tools

1. Photograph where located.
2. Protect working ends.

Packaging Plastic bags.

Significance Location of paint on cutting edge which may match paint at the scene and to link the tool to a particular toolmark.

Toolmark

1. Photograph (overview, midrange and close-up with scale).
2. Make a cast.
3. Recover complete item for further examination if it is portable.

Packaging Rigid plastic container or plastic bag.

Significance Link toolmark to a particular tool.

Vegetation

1. Photograph various types of vegetation.
2. Collect samples consisting of complete plants and roots.

Packaging Paper bags with cardboard stiffening to prevent damage.

Significance Identify species and compare with trace material found on suspect or his or her environment.

Wires

1. Photograph site.
2. Protect ends of wire.
3. Label ends cut by crime scene investigator.

Packaging Plastic bags.

Significance Identify tool type and compare with tools submitted for examination for possible identification of the tool.

Notes

1. Potential evidence should be recovered and submitted to the laboratory as soon as possible and examination of individual items at the scene should be kept to a minimum.
2. The above are *general* guidelines. Different jurisdictions may advocate some variation on these procedures.

Preservation

H B Baldwin, Forensic Enterprises Inc., Orland Park, IL, USA
C Puskarich May, Criminal Justice Institute, Little Rock, AR, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0453

Introduction

For the purpose of this article, ‘preservation’ in the realm of law enforcement will involve two areas: preservation of the crime scene and physical evidence preservation. The primary purpose of both forms of preservation is to limit or eliminate the potential for contamination or destruction. In doing so, the crime scene/physical evidence can be appropriately processed and documented. Securing and protecting the crime scene will lead to a more accurate reconstruction of the crime and have an impact on its solvability.

Criminal Activity Time Lines

With any criminal activity, there is always a time line. Before the criminal act, several things have occurred. If the crime scene is a private location, such as a residence, the activity is limited to the occupants; if the crime scene is a public place, such as a school, business or any place of common access, the area can and usually will be frequented by a larger number of persons. The time line continues with the actual criminal act. The activity of the suspect(s) will vary depending on witnesses/victims present, whether it is a public or private location and the type of crime committed. The scenes and evidence will be in a state of fluctuation. Once the criminal act has been completed, the suspect(s) flees, creating additional destruction and possibly contamination of the scene and evidence. As the suspect(s) flees, evidence may be dropped, intentionally or accidentally, creating an area secondary to that of the original crime scene.

These scenarios will be altered by the time between the suspect(s) fleeing and the arrival of the first responder. In the case of residential burglary, the crime may go unnoticed for several hours, or possibly days. This expanded time frame can greatly influence the quality and value of the evidence. If the scene is located at an airport, for example, it could be totally destroyed by the chaos that would probably occur.

First Responders

Once the first responding officer arrives, the crime scene will be under control. Or will it? To preserve the scene the officer must take control of it. The size and nature of the crime will determine how difficult this task may be. If the scene is a residential burglary, then the officer may have a much easier job, requiring possibly only his or her presence to preserve it. However, most dwellers want to know what was missing, and either will have gone through everything before the officer arrived or will go past the officer to see if that favorite piece of jewelry is missing. In either case the scene becomes contaminated, or at least the evidence potential of the scene is minimized. In the airport scenario, the preservation of the scene becomes difficult and is magnified by the size of the scene and the number of people present.

The first responding officer must gather information and complete several tasks upon arriving at the scene of a crime. He or she needs also to check the premises for the presence of the suspect(s), determine the layout of the building or area, conduct a cursory examination of a hazardous situation, provide emergency aid to the victim(s), examine for evidential potential, protect ingress and egress areas and identify temporary evidence. While doing all this the officer must also preserve the scene from further contamination and potential evidential losses.

Dimensions of the Crime Scene

As additional assistance arrives, the security of the scene becomes even more critical. Identifying the dimensions of the scene is now essential. All crime scenes have two areas: the primary and the secondary. The primary area is the immediate area of the crime; the secondary area includes all avenues leading to, and away from the scene. Once this area is defined, barrier tape can then be placed to identify the physical boundaries of the crime scene. Whether barrier tape, wooden barricades or any other form of barrier is used, the physical barrier identifies the limits of the scene and determines how closely individuals can actually come to the scene. Unfortunately, crowds may appear before the limits of the scene are determined. If this occurs, potential evidence in the secondary area(s) of the crime scene may be trampled, destroyed and contaminated. Consequently, the limits of the scene must be addressed immediately, especially at a major crime scene.

The size of the outer perimeter of the crime scene will determine the number of officers needed to secure it. Consider, for example, a bank robbery. The perimeter will depend on the location of the bank. Is it on a corner or in the middle of the block? Is it a stand-alone building or nestled in with other commercial buildings? How does securing the perimeter change if the crime occurs at an airport? Actions to be taken to secure the perimeter will vary greatly, depending on where the crime occurs.

The scene must also be protected from other law enforcement personnel. The common practice now is to identify an inner perimeter (or cordon) with 'biohazardous' banner tape, and the outer perimeter (or cordon) with 'crime scene – keep out' banner tape. Securing the scene in this manner allows supervisors and other officers to come closer to the scene but identify the area that is restricted.

Crime Scene Logs

Personnel logs are required at every crime scene. Minimally, who was there and when must be recorded. This information can be crucial in determining what standards may be required for comparison with potential evidence found at the scene. In a familiar situation, a paramedic has administered first aid to the victim at a crime scene and then transported the victim to the local trauma center. A bloody footwear impression is recovered by crime scene personnel and

thought to be the suspect's shoe print. The crime scene log should provide a listing of personnel at the scene. This information can then be used to eliminate or include individuals as potential depositors of the footwear impression found at the scene. Similar standards would also be required for major case prints, biological evidence, trace material and hair. The more evidence available and the greater the number of personnel in the secured areas of the scene, the more work is created for the forensic analysts and the longer the delay in the effect this evidence may have on the investigation.

A major crime scene log will typically include the victim(s), witness(s), first responding officer, paramedic or fireman, investigator(s), crime scene personnel, supervisors, coroner or medical examiner, and possibly the prosecuting attorney. The log should accurately record those who entered and exited the scene by identifying them by name and agency; why they entered; and the times of entering and leaving the scene. A time line of the actions at the crime scene is also created through the information available in crime scene logs.

Command Posts

Command posts (or rendezvous, RV, points) are normally used at large scenes or those of notoriety; however, a form of a command post should be used at every scene. The command post is simply a staging area where information may be obtained on what has happened at the scene. It is also used for planning and decision-making activities for the crime scene. At a small scene it may very well be the kitchen table; at a large scene it will be a custom-made mobile command vehicle specifically designed for communication, by phone, fax or video, with the main headquarters or the news media. In larger scenes the command post is normally in the outer perimeter region of the crime scene.

Crowd Control

Crowd control at the crime scene is normally accomplished by the use of a physical barrier. This barrier identifies to all the limitations to access. If the crowd becomes unruly then additional assistance is required to control their access to the crime scene area. Evidence will be destroyed if the scene is not controlled.

Law Enforcement Control

Once a scene has been secured from the public, what about the law enforcement personnel and others present at the scene? The banner tape is effective, but may not stop an official from entering the scene. Written policies must be available in the law enforcement agency to identify the protocol at a crime scene. Officers will understand the ramifications of violating a departmental order against crossing a 'police line' in such a scene. The policy must also dictate who is in charge at the scene and delineate his or her authority.

Outdoor Scenes

Outdoor crime scenes are difficult to control for many reasons. Besides trying to determine the dimensions of the scene and providing sufficient security from onlookers and the media, the other major problem is that all is in view, to be photographed and watched by the news media, who telecast the events live. This hampers law enforcement, in the sense that there are some items of evidence that may not be for public information. In recent years portable privacy fences have been used by law enforcement agencies to preserve the scene from prying eyes. There is also the issue of modesty for the victim and the family. Privacy screens preserve the integrity of the scene and modesty of the victim.

Scene Conclusion

Once the extent of a crime scene has been established, it can be secured and preserved for the processing and collection of physical evidence. This preservation will vary from scene to scene and must be enacted by the responsible law enforcement agency as quickly as possible. The scene needs to be protected not only from 'outsiders' but also from the police themselves. Most crime scenes require that the first responding officer be sufficiently equipped to secure the scene temporarily by using banner guard tape to identify the dimensions of the scene. All other responding officers and personnel can then see the boundaries and move within them accordingly. This small act alone can be the most valuable asset to all crime scenes. If the scene is secured quickly, then the likelihood of solving the case is greatly increased.

Preserving Physical Evidence

The type of physical evidence at the crime scene can present logistical problems. If the evidence is temporary – for example, footwear impressions in snow, blood patterns on asphalt on a hot day, etc. – this must be addressed first by the officers. The preservation of this type of evidence is usually completed first at the crime scene once the general area has been documented. Both of these examples can be preserved until processing by simply placing a covering over

them to prevent further contamination or destruction.

The processing of the evidence at the crime scene requires the use of procedures that reduce its destruction. All evidence should be first visually examined with the naked eye, then examined with the use of a forensic light source, and then processed in a nondestructive manner. Each of these steps must be performed carefully, in such a manner that the potential value of the evidence to the investigation is not diminished.

All physical evidence, once documented and processed at the scene, must be packaged properly. Packaging can be critical to its protection. If an item with a nonporous surface, such as a drinking glass, is packaged to be transported to the police agency or laboratory for further analysis, it must be packaged in a paper product so it can 'breathe' and not become moldy. Packaging must also be done in a manner that preserves potential latent prints. For example, if the paper bag is too large, the glass could rub up against its walls and 'wipe' the prints. The packaging must be appropriate for the evidence.

An essential element of the preservation of evidence is the sealing of the evidence bags or containers. In order for the evidence to be viable for court, its packaging must be properly marked and sealed. All evidence must be placed in a sealed container before it leaves the crime scene. This is usually accomplished using integrity (tamper-evident) tape. The tape should also have the initials or marks of the person recovering the evidence placed across it. A new style of integrity tape now available includes a space for the packer's right thumb print. This combination of integrity tape and markings will ensure the evidence has been satisfactorily protected from tampering.

Preservation of the evidence does not stop with packaging. The evidence must be handled properly and stored in a temperature-controlled environment. Leaving items in a hot vehicle for long periods of time could invalidate some analyses. Temperature, whether it is extreme heat or freezing, can greatly affect outcome.

Conclusion

Preservation techniques must be used in every aspect of crime scene processing and physical evidence collection. To prevent the loss or destruction of the physical evidence, the scene must be secured in a timely fashion and physical barriers must be used. Physical evidence is normally fragile and therefore requires that great care is used in its handling. By doing so, the investigative value of the physical evidence recovered from the crime scene can be maximized.

See also: **Crime-scene Investigation and Examination**: Recording; Packaging; Contamination; Major Incident Scene Management.

Further Reading

Eliopulos LN (1993) *Death Investigator's Handbook: A Guide to Crime Scene Processing, Forensic Evaluations and Investigative Techniques*. Boulder, CO: Paladin Press.

Fisher BAJ (2000) *Techniques of Crime Scene Investigation*, 6th edn. Boca Raton, FL: CRC Press.

Hawthorne MR (1999) *First Unit Responder: A Guide to Physical Evidence Collection for Patrol Officers*. Boca Raton, FL: CRC Press.

US Department of Justice, National Commission on the Future of DNA Evidence (1999) *What Every Law Enforcement Officer Should Know About DNA Evidence*. Washington, DC: Department of Justice.

Recording

J Horswell, Australian Federal Police, Canberra, Australia

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0449

Background

The accurate recording of details of a crime scene, incident scene or of any subsequent examination is important for several reasons. It is important for the crime scene investigator as it will provide the basis for statements and reports that the crime scene investigator has to compile at a later date. It will also provide investigators with information of which they may not otherwise have knowledge; it will assist the court in reconstructing the scene and may provide the most reliable facts regarding the case. Finally, it may provide the courts with the best evidence available.

It is fair to say that crime scene investigators can never make too many notes during a scene investigation. Notes should always be compiled during the course of the examination, not at some time later; however, if it is not possible to do so then details should be recorded as soon as possible afterwards.

There are obvious and very good reasons for compiling contemporaneous and accurate notes:

- Notes made at the time of an examination are likely to be more reliable and accurate than notes made some time later.

- By making notes as he or she is working, the crime scene investigator is less likely to overlook minor details committed to memory.
- An accurate record of times and dates will be maintained. This will avoid discrepancies with the records of other investigators involved in the investigation.
- An accurate record is available for later reference during the investigation and when compiling statements for court.
- When giving evidence in court, the crime scene investigator may be permitted by the court to refresh his or her memory by referring to the notes taken during the investigation.

Obviously if the notes are made during the conduct of each stage of an examination, then there can be no dispute as to their recency.

Photographs

Photographs can provide a detailed record of the condition of a scene, illustrating the items present and their relative locations. For this reason, photographs should be taken before items are moved or interfered with, and should be taken from varying angles.

There may be items shown in the photographs which were not mentioned in the written notes taken at the time, and the photographs may help to refresh the crime scene investigator's memory on some aspect of the scene or examination. On the other hand, during court hearings, defense counsel may cross-examine the crime scene investigator about objects shown in the photographs, and if there are no notes about the items it may prove embarrassing. It is therefore important not to rely too heavily on photographs: the crime scene investigator should insure there are ample written notes as well.

The general survey procedure will help determine what photographs will be required and the sequence in which they are taken. As a rule, the crime scene should be photographed after the general survey and before any further examination is made, without reconstructing the scene in any way. The crime scene investigator should be able to demonstrate photographically how the scene was before the start of the scene examination.

It is not the intent of this article to provide a short course on photography; however, the points raised will insure that there is an adequate coverage of the scene by the crime scene investigator.

The $6 \times 6/6 \times 7$ cm or 120 roll film camera is ideal for the recording of toolmarks, shoe impressions and tire tracks. The Mamiya camera is a popular large format camera; these cameras are heavy and therefore better used with a tripod. This is also an excellent format for use during a postmortem examination for the recording of injuries.

The 35 mm camera is the workhorse of the crime scene investigator: a good, all-round camera for scene work and the recording of fingerprints.

Before commencing the photographic aspect of the crime scene investigation it must be remembered that photographs should not include artifacts introduced by the crime scene investigator and other investigators. Briefcases, clipboards, photographic equipment bags, crime scene kits or the photographer's feet should not feature in any of the photographs.

Each crime scene will be different but the following should be considered:

- The photographic record should be comprehensive and should include the general layout of premises or features of an area. This will depend on the seriousness and circumstances of the crime.
- The photographic record should illustrate the relative position of rooms, the state of those rooms and the position of houses in streets in relation to the crime scene.
- Footprints, tire tracks and toolmarks should be photographed with a scale before casting. A close-up and positioning photograph should be taken.
- Photographs should be taken from a number of angles or positions, including those described by witnesses.
- A series of photographs should be taken from the point of entry to the point of exit.
- Detailed photographs should be taken of potential evidentiary material, such as the body, injuries, weapons, trace material, cartridge case/s, damage and other relevant items.
- As the scene examination progresses, further photographs should include new potential evidentiary material found, or areas of importance which were previously concealed.

Before taking any photographs the crime scene investigator must think:

- What am I going to photograph?
- Why should it be photographed?
- What do I want to demonstrate using photography?
- How can I record it as I see it?

Having made these comments, it is necessary to cover all pertinent material. It is wiser to take too many photographs than too few. It must, however, be remembered that it is not necessary to have all the negatives printed. This should create no problem in court as long as the investigating officer is aware of

the situation. It may also be necessary in some jurisdictions to advise the defense of their existence; one way to cover this point is to have the investigating officer select the photographs he or she wants from proof sheets. The crime scene investigator should be prepared to defend his or her selection of the photographs later in court.

Video Recording

It is useful to video the crime scene; a recording may be an invaluable briefing tool for investigators and others to view later, as well as to introduce as potential evidence in court.

The recording of a crime scene by video should be undertaken in each serious and major crime. Experience has shown that the video of any crime scene should be taken without sound. The subsequent audiences that view the video should be guided through it by the crime scene investigator or investigating officer, either in person or by means of a 'voice over'.

The video recording of what is called a 're-enactment' should be attempted only after the suspect has been interviewed and only after an invitation is accepted by the suspect, with the video being taken while the suspect is under caution. Such videos have been shown to be a very successful tool in presenting the prosecution case at court. The court will also be able to see if the suspect is under stress or duress at the time of the re-enactment video, along with his or her general demeanour and that of the interviewing officer. Experience has shown that powerful evidence can be gained from this technique.

The video recording of a crime scene should be under the direct control and guidance of the crime scene investigator or crime scene manager, as it is only these individuals who are aware of the current position regarding the processing, by recording, search and recovery, of potential evidentiary material at the crime scene.

Digital Photography

Digital photography is not currently popular as an alternative to silver halide images as the image can be altered. However, many systems are being developed with an inbuilt 'audit trail' to overcome this problem. There is every likelihood that digital imaging will overtake the more traditional recording methods in the future.

This medium is also a useful investigative tool that can facilitate the movement or shipment of an image of an item from one place to another as an attachment to an e-mail message. It may also be used in the recording and movement of arrested person photographs. It may also come into its own in the presentation of evidence in court. Courts are already using CD-ROM technology as a means of presenting evidence. This includes statements, photographs and drawings. In addition, virtual reality crime scene footage is also being used to present the crime scene to the court. Crime scene investigators should be aware of its potential as an aid to a forensic investigation.

Labels: The label should be completed at the time of collection or receipt of the item so that the item can be later identified, as well as to start the chain of custody for the item.

Written notes: As mentioned above, it is essential for the crime scene investigator to make notes as he or she progresses through an examination, or as soon as possible afterwards. Because of this, the notes may not be in a logical sequence, but this does not matter. The sequence can be reorganized at a later stage, when writing the report or statement.

The main aim of writing notes is to provide an accurate and comprehensive record of events and observations which will still be meaningful months later. For this reason, it is preferable to write detailed notes at the time rather than attempting to save time by using abbreviations, which, although readily understood when being written, might be insufficient to refresh the crime scene investigator's memory after several months.

On arrival at a crime scene, the following should be noted:

- date and time of arrival;
- names of persons at the scene on arrival;
- weather conditions;
- lighting conditions at night;
- what has happened – the incident;
- what has taken place – activity since incident;
- officer in charge of the case;
- scene guard;
- assistance provided at the scene;
- other resources already requested.

The sequence of the crime scene investigator's actions following arrival will vary depending upon the situation with which he or she is faced. If there is no requirement to start a particular examination immediately, it is often advantageous to spend some time studying the crime scene, noting all observations. Any movement through the crime scene, noting observations, can only be done if there is no risk of contaminating or damaging possible evidence. A pathway should be identified, which is used as a common approach path into and out of the critical areas of the crime scene.

Plans

There are two types of plan: a sketch drawn by the crime scene investigator, and a scale plan, which can be drawn by an experienced crime scene investigator or a draughtsman. These complement written notes and photographs and are notes of the scene examination as well.

Sketch plan

A sketch enables the crime scene investigator to show the location of items and their relationship to other items. A sketch should be drawn for all serious and major crime scenes.

Although the sketch only needs to be freehand, it must be neat enough for the crime scene investigator or draughtsman accurately to interpret the data at a later date in order to produce a scale drawing.

There are several basic types of drawing that are commonly encountered in sketching crime scenes. The floor plan view is the most common and is the easiest to complete. It depicts the location looking down from above. This should be used for both indoor and outdoor scenes. The exploded view or crossprojection method is similar to the floor plan view and differs only in that the walls fold out to reveal items of evidence found on or in the walls. Isometric projection of walls as stand-alone drawings may be used to indicate items of evidence, such as bloodstain patterns found on walls at a crime scene, exhibiting extreme violence. Three-dimensional drawings, virtual reality and animated computer programs are now being used more and more in crime scene investigation.

The following two methods are suitable for measuring crime scenes:

- *Coordinate method:* This method uses the principles of measuring the distance of an object, such as a body, from two fixed points. One form of the coordinate method involves the use of a baseline, which is drawn between two known points. The baseline may also be a wall or drawn as a mathematical centre of a room, the exact dimensions of which are known. The measurements of a given item are then taken from left to right along the baseline to a point at right angles to the item which is to be plotted.
- *Triangulation method:* The triangulation method requires three measurements:
  - base;
  - shortest side of the triangle;
  - longest side of the triangle.

An established base may be used, for example the side of a house. Two measurements are then taken, from the corners of that side of the house to the item to be plotted. When a crime scene is in an open area, such as a beach, paddock or park, the triangulation method is usually employed but it is necessary to establish a base. This can be achieved with the aid of a magnetic compass to determine true north.

Method of measuring crime scenes

- Accurately determine north with a compass and place it at the top of the plan.
- Determine what is to be included in the plan and the method of recording.
- Draw a rough sketch on which the measurements will be recorded.
- Work systematically throughout the scene, recording dimensions, in the case of a room, and the location of important items within it.
- It is ideal that the person responsible for both the sketch plan and the scale drawing should be the person who records the measurements on the sketch plan.
- Use the progressive system of measurement where possible; for example, corner of room to the nearest point of window 0.3 m metres and 3.5 m to the other end of the window frame.
- In order to locate items within a room or open area, use either the coordinate or the triangulation method, or a combination of both.
- The position of bodies and important items should be plotted prior to removal or collection; however, the position of 'fixed' objects may be recorded at a subsequent date, thus enabling a quicker examination of the scene.
- If objects must be moved prior to plotting then mark their location before moving them, for example with chalk, felt marking pen, crayon or spray paint.
- Add the crime scene investigator's name, the case, date, time and location. If anyone assisted, his or her name should also be included on the sketch.

Scale plan

Scale plans are used to convey accurately the size, shape and position of important potential evidence and other features of the crime scene. They are a valuable adjunct to scene photographs. Scale plans are also an aid in reviewing a crime scene with investigators.

The use of modern surveying equipment overcomes many of the problems encountered in preparing crime scene plans. These tools are now being put to good effect by many practitioners.

Summary

The fundamental reason for recording crime scenes is to take a crime scene and reproduce what has occurred for the information of the investigation team and, ultimately, the court.

See also: **Crime-scene Investigation and Examination**: Major Incident Scene Management. **Photography and Digital Imaging**: Digital Imaging Enhancement.

Further Reading

Fisher BAJ (1993) *Techniques of Crime Scene Investigation*, 5th edn. Boca Raton FL: CRC Press.

Goddard KW (1977) *Crime Scene Investigation*. Reston, VA: Reston.

Horswell J (1993) Crime scene investigation. In: Freckelton I, Selby J (eds) *Expert Evidence*. North Ryde, NSW: Law Book Company.

Kirk P (1974) *Crime Investigation*, 2nd edn. Chichester: Wiley.

Svensson A and Wendel O (1974) *Techniques of Crime Scene Investigation*, 2nd edn. New York: Elsevier.

Recovery of Human Remains

H B Baldwin, Forensic Enterprises, Orland Park, IL, USA

C Puskarich May, Criminal Justice Institute, Little Rock, AR, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0451

Introduction

This article will introduce the reader to the basic techniques used in the recovery of human remains. It is not intended to be all inclusive, but may provide insight into unique aspects of recovering remains suspected to be human. Physical evidence, testimonial evidence, search techniques, surface recoveries, excavations and exhumations will be discussed. Recommendations to maximize the amount of information obtainable by a forensic anthropologist will also be presented.

Recovering human remains that are buried and/or badly decomposed or skeletonized is the same as any other death investigation, with the four basic types of manner of death (homicide, suicide, accidental and natural) possible. The only difference between 'normal' death investigations and recovery scenes is the amount of decomposition, the disarticulation of the bones, and the unusual environments within which we must work. As in all death investigations, recovery of human remains is a team effort. The combined talents of several people, each a specialist in his or her own area of expertise, is required. All law enforcement officials involved have specific responsibilities. Meeting these responsibilities often impacts the efforts of others.

Although a crime scene can be handled in numerous ways, and each case may dictate a different approach, experience has proved certain guidelines to be advantageous. The following are suggestions for consideration in any death investigation.

First officer on the scene

- Render aid if needed.
- Protect the scene; this includes prevention of possible contamination by keeping other police personnel from entering the crime scene.
- Secure witnesses and keep them separated.
- Take immediate notes because the scene will continually change.
- Request the services of a crime scene technician, the coroner and investigators. If required by policy, notify the state's attorney's office.
- Condense all facts into a comprehensive report.

Crime scene technician

- Locate and identify physical evidence at the scene.
- Process the crime scene, attend the autopsy and take custody of all physical evidence.
- Package and transport all physical evidence to the appropriate crime laboratory.
- Prepare a comprehensive report.

Coroner/medical examiner

- Give direction to the pathologist.
- Determine the cause and manner of death. Request assistance from the crime scene technician if needed.
- Establish a close line of communication between the coroner's office, crime scene personnel and investigators.

Pathologist

- Obtain all physical evidence from the victim's body. If all pertinent information and evidence are recovered from the body, there should be no need to exhume the body. Items such as a knife tip broken off inside the body can be used for laboratory comparison if the knife is recovered.

- Communicate with the coroner and the crime scene personnel at the autopsy.
- Consult with a forensic anthropologist.

Investigator

- Gather all possible information from the first officer on the scene, any witnesses, the crime scene technician, the pathologist, the anthropologist and the coroner.
- Authorize one person to release information and maintain continuity throughout the investigation. Information released without thought to investigative leads or case prosecution may jeopardize successful resolution and prosecution of the case.
- With the state's attorney's (or prosecutor's) office, determine the value of lead information and the direction of the investigation. The state's attorneys office and the investigator will be charged with the long-term aspect of the investigation and the ultimate prosecution of the case.

Thorough investigation of a crime must be a joint effort, otherwise the citizens are underserved. If each individual completes his or her portion of the process, the outcome will yield the most effective results.

Scene Processing Procedures

In processing a 'normal' death scene, primary and secondary areas of the crime scene are identified. The primary area represents the location of the body, while the secondary area is the area(s) leading to the crime scene. For instance, if a body is found in the middle of a room, the primary area of the crime scene is in that room and is consequently well defined by the walls, floors and ceiling. The secondary area represents all avenues to the room, including the outside yard. How did the suspect(s) enter and leave the scene? Did he or she drive to the scene? Answers to these questions define the secondary areas of the crime scene.

If the crime scene is an outdoor scene with a body lying in an open field, what are the dimensions of the scene? Answering this question is problematic to most. The primary scene is represented by the area closest to the body, while the secondary scene is the surrounding access area. The body didn't just fall out of the sky and land there. Someone used a vehicle to transport the body there, or carried it to that spot. Therefore, prime physical evidence leading up to the body, e.g. tire tracks, footwear prints, drag marks, etc., may be present.

Unfortunately, whenever there is a body in a field or any open space, walking in a straight line right to the body, ignoring the evidence on the way, is a natural tendency. We seem to lose sight of the dimensions to the scene. To correct this tendency, identify a least likely point of entry to the body and approach this way. Open your eyes and look for evidence in the path the perpetrator most likely used, it's going to be there!

Two types of evidence, physical and testimonial, are possible at all crime scenes. In general, physical evidence is anything that can be held; testimonial evidence is what can be testified to in the reconstruction of the crime scene. One form of evidence is just as important as the other.

Death investigation scenes are processed to prove whether a crime was committed and provide clues to who may have committed the crime. Determining the identity of the deceased is an added problem that arises in cases involving human remains which are buried and/or badly decomposed or skeletonized. Current procedures for identifying the deceased are by finger-, palm- or footprints, DNA, dental charts and X-rays, comparison X-rays of old injuries and deformities of the deceased, measurement of the bones of the skeleton to determine approximate age and height, existence of physical deformities, such as missing fingers, etc., and tattoos. Facial reconstruction by a forensic sculptor/artist or an anthropologist can also be attempted. A likeness of the victim is created and circulated in local, state and regional newspapers. This likeness may be recognized and open new leads as to the identity of the victim. All new leads are then confirmed using the traditional methods of positive identification discussed above. The victim's personal identification might also be found on the body. Even in this situation, the identity of the individual must be confirmed using traditional identification methodology.

What are the normal procedures in processing a death scene? The following chain of procedures is recommended:

1. *Interview:* gather and verify information.
2. *Examine*: examine and evaluate the crime scene.
3. *Photograph:* photograph the crime scene.
4. *Sketch*: sketch the crime scene.
5. *Process:* process the crime scene.

A scene that involves the recovery of buried and/or decomposed or skeletonized human remains is no different from any other crime scene investigation. There is still a scene, a victim and a suspect. As with all other types of scenes, examining and processing the scene for physical evidence is required.

What is physical evidence? The best definition available is:

> Wherever he steps, whatever he touches, whatever he leaves, even unconsciously, will serve as silent witness against him. Not only his fingerprints or his footprints, but his hair, the fibers from his clothes, the glass he breaks, the tool mark he leaves, the paint he scratches, the blood or semen he deposits or collects – all of these and more bear mute witness against him. This is evidence that does not forget. It is not confused by the excitement of the moment. It is not absent because human witnesses are, it is factual evidence, physical evidence cannot be wrong; it cannot perjure itself; it cannot be wholly absent, only its interpretation can err. Only human failure to find it, study and understand it, can diminish its value. (Harris v. United States, 331 US 145, 1947)

The only thing that has changed since 1947 is the interpretation of the evidence and the ability of forensic science technology to identify potential evidence. Today, more in-depth analysis of the evidence is possible. In addition, crime scene investigators are better educated and trained to collect and preserve evidence and have more advanced equipment, which increases the likelihood of identifying and recovering significant physical evidence.

What types of physical evidence can be found at a recovery site? The possibilities are endless. Factors that can change or diminish the value of the evidence are time since death, weather conditions and contamination by others (including investigators).

Surface recoveries and excavations are processed in the same manner as any other crime scene. Physical evidence is present and must be identified. The same processing techniques are used: interview, examine, photograph, sketch and process. There is no reason to rush and do a sloppy or substandard job. This is one of the few cases in which time is on your side!

Search Techniques for Human Remains

The primary event in the recovery of human remains is finding the body. Most remains are found by accident. In the Midwest and South of the United States a majority of the remains are found in the fall or spring by hunters, farmers working their fields, or others such as hikers in the woods and construction workers performing excavations at job sites. Sometimes remains will be found incident to searching an area specified by investigative leads or informants. Once the remains are found, define the primary area of recovery. Don't forget the secondary area(s)!

Prior to searching an area, the information received must be verified. Nothing will make supervisors unhappier than going on a wild goose chase, depleting financial and human resources. Wild goose chases are also emotionally difficult for the family of the victim.

Once the search area is defined and the information verified, personnel and equipment needs must be evaluated. A team must be established. A 'team leader', searchers and crime scene specialists to handle the photography, sketching, collection of evidence and the recovery of the remains are key team members. A detailed map of the area in question and aerial photos to show the layout of the search scene may be available and should be obtained. It is suggested that someone familiar with the terrain also be contacted to provide an insight into problems which may be encountered, such as identifying any changes made to the area which are not documented.

Comfort conditions must also be taken into account by the person in charge of the search or recovery site. Weather conditions and provision of food for the workers, bathroom facilities, as well as liquids to drink and water for cleaning up, must be considered in planning decisions.

Once the area has been evaluated and a team is established, the object of the search and how to look for it must be communicated to the team. Do not assume that they know what to do. Have a meeting prior to the search and explain in detail the objective of the search.

Normal field search methods are used for all outdoor crime scenes. The following four methods, or a variation of them, are used in searching outdoor areas for the recovery of human remains:

- *Grid* The area is sectioned off in a grid with the searchers in one area of the grid at a time. This method is normally used when the area is quite large, such as several acres of land. Each area of the grid is then searched by circle, strip or zone.
- *Circle* This consists of a center stake with a rope tied to it, with the searchers at 1.5 m intervals circling around the center in a straight line. This method is not frequently used but can be effective with large areas and several searchers.
- *Strip* This consists of a line of searchers approximately at arm's length away from each other in a straight line walking through the area to be searched. Using barrier tape to define the strips will help.
- *Zone* This is similar to the grid search except that it is normally used in smaller areas to be searched or a portion of the grid. The zone is then searched with the strip method.

The method selected will depend entirely on the terrain to be searched and the number of people available. The area should always be searched twice but not by the same person. A second person may find what the first person missed.

When using any of the search techniques, if one of the searchers finds an item, the whole team should

stop until the item found can be photographed, the location sketched, and the item collected and marked for further examination later. After the item is marked or collected the searchers can continue until another item is located. This process is continued until the search of the entire area is completed.

Another method to consider is use of trained dogs. Some dogs are trained for recovering items of evidence in the field, while others are trained for finding deceased bodies. Different methods are used in the training of these dogs. Check with your local canine unit to ascertain if they can perform either function. Only dogs specifically trained for the search and recovery of human remains should be used. Dogs crosstrained for drugs, arsons, bombs and other items of evidence are not as productive as those specifically trained for recovery sites.

Looking for buried remains requires the use of the same search methods as for surface recoveries, but different search indicators are utilized. As a result, it is imperative that searchers are educated as to the different types of possible burial indicators. Because of time and weather elements, burial indicators may be difficult to find. Which indicators are expected will vary depending upon the type of terrain being searched. Remember that you may be looking for pieces of a body, which could mean multiple burial sites, rather than one large burial site.

Disturbed vegetation, soil compaction and soil disturbance are some of the indicators of a burial site. Multiple indicators may be present at any site. These indicators are discussed below.

Disturbed vegetation Whenever a hole is dug in the ground, the vegetation in and around the hole is disturbed. This vegetation will no longer remain as it originally was. It may be upside down, with the roots showing, or just brown from being uprooted. Adjoining areas, disturbed during digging, will also show signs of vegetation disturbance. In a wooded area, a small clearing in the trees may indicate a burial site.

Soil compaction The natural decomposition of the buried remains will leave a void in the soil. Through time and rain the soil above the remains will sink to fill the void, thus forming a depression in the surface above the body. This depression is sometimes called a compaction site. A secondary depression may also be noted inside the primary depression. This is caused by the abdominal cavity deteriorating. Again depending on time and weather factors, this depression may have new vegetation or even trees growing from it. New vegetation may also be a burial indicator because the new growth will not be as mature as growth in the surrounding area. There may also be visible cracks in the soil outlining the actual grave. These cracks are made when the disturbed soil shrinks from the undisturbed soil owing to the depression in the ground.

Animals can cause similar compaction by burrowing holes for nests or looking for food. Other decomposing material will also cause the same depressions. However, an area of compaction of approximately $0.5 \times 2\,m$ will usually indicate something large is buried.

Soil disturbance When a grave is dug, the layers of the soil are disturbed. The soil under the ground is layered. Some areas will have very shallow layers or multiple layers within a few centimeters from the surface, while others will have layers meters thick. At different depths, the soil will vary in color. These different colors represent the different layers of soil. For instance, black soil might be found from the surface to a depth of about 10 cm, a lighter color of soil might follow for several centimeters, with clay below. All these layers may repeat themselves or be in different orders. However, once the layers are disturbed no amount of effort and precision can replace them exactly the way mother nature put them there.

Digging not only disturbs the soil layers in the grave, but also disturbs the surface soil around the grave. There will always be some residue left after refilling a hole. The residue will be a different color from the surrounding surface soil.

Some special problem areas for visual indicators are sandy beaches, desert areas and cultivated land. Visual indicators on the surface may be difficult to find; therefore, locating a burial site in these areas will require a different approach to the problem. Several methods are currently being explored, as outlined below.

Infrared photography

This method uses the difference in temperature between the buried body and the temperature of the soil around or on top of it. Infrared photography may also indicate the difference in temperature of disturbed and undisturbed soil. Success with this method may be minimal, depending upon the length of time since the event.

Methane detector

Any organic object that is decomposing will produce methane gases. Unfortunately, the decomposing item may be a tree stump, vegetation, an animal or garbage. In addition, the longer the body has been buried, the less likely it is that methane gas can be detected. This method is therefore seldom used in the field.

Aerial photography

The comparison of aerial photographs of the suspected area taken over a period of years might disclose a possible burial site. Aerial photos could show a vegetation disturbance occurring where a body is buried. These photographs are usually taken for tax assessment purposes and are available for several years. The aerial photographs will also show what new construction has taken place during the period in question.

Probe

In using this method a four foot metal rod approximately 1 cm in diameter with a 30 cm handle forming a 'T' is poked into the ground. In pushing the probe into the ground, a difference in the pressure needed to push the probe into undisturbed and disturbed soil should be felt. To be more confident in detecting these differences, extensive practice should be undertaken before actually probing the suspected burial site. A 'feel' for this must be acquired by sampling the ground in the area. When using the probe, sample holes should be obtained every 30 cm. This method requires the searcher to use *extreme* caution to avoid potentially damaging the body and other possible physical evidence. To avoid such damage, a minimal amount of pressure should be used on the probe at all times. Feeling an increased resistance will be a strong indication to proceed with extreme caution or retrieve the probe. The amount of time needed to probe an area appropriately and the subsequent wear on the searcher are other drawbacks to this method.

A variation of this method is to use a similar probe with a larger, hollow shaft for taking soil samples. These soil samples are then tested for the presence of calcium (bone) that leaches into the soil from the decomposed body. However, this type of testing cannot be done on site. Samples must be sent to the laboratory for testing; therefore, if time is of the essence, this would not be the method of choice.

Construction equipment

Construction equipment should be used as a last resort, but some situations may call for it. Use experienced backhoe and bulldozer operators who are capable of removing layers of 5–30 cm at a time. When carefully done, soil disturbances, or the grave itself, may be identified. More damage to the remains can be done by this method than any other and it should therefore be used only when no other options are available.

Some of the search techniques discussed above require good old-fashioned manual labor. The best technique to use will depend on the terrain, the size of the area to be searched and the resources available.

Surface Recovery

Once the search is completed and the body located, the recovery site must be defined. Extreme scattering of the bones/body parts or physical evidence by animals frequently occurs. Therefore, the area encompassing the scattered bones may range from a short distance to several meters. Some of the bones may never be found because of vast scattering or consumption by animals. Depending upon the time of year as well as length of time since death, the bones may even be covered by vegetation, dead leaves or fallen trees or branches. This covering of the deceased may also have been done intentionally by the suspect to camouflage the body.

Before approaching the remains, a general indication of the probable path of the perpetrator must be established. Do not take this path. Rather, avoid this path at all costs. Not doing so may cause valuable evidence to be destroyed or contaminated.

Once the remains have been located and the recovery area defined, proceed as with any other crime scene: secure the area, examine and evaluate the site, photograph, sketch and process. An evidence-free access to and from the site (i.e. corridor) must be established prior to processing. A metal detector should be used before declaring this area clear of evidence. In addition, an outer perimeter search, using a method described in the previous section, must be completed to locate other body parts or physical evidence. A command post should be set up, preferably away from the recovery site. A checkpoint should also be set up to check personnel into and out of the scene and to limit the number of people who are entering the site. One way of limiting the trouble caused by other people entering the scene is to take Polaroid pictures of the site and leave them at the command post for viewing.

The most difficult part of this recovery is now over. Take your time and do not permit anyone to rush you. Do it right the first time because there are no second chances.

After the site is photographed and sketched, clear away all the vegetation and debris. This must be performed in a way that avoids disturbing the remains or any of the physical evidence. Photos should be taken of the new 'clean' site. Using rope or string, a grid should be set up for the purpose of locating the items by measurements and for ease in placing the items on a sketch. The grid should be measured so that the sides are square to each other. A metal detector should again be used before any further processing. Any items located should be marked with a

wood or plastic stake for future reference. Plot all evidence and remains on the sketch. Closeup photographs should be taken of all items before their removal. All photographs should be taken with a scale to show the size of the item.

All evidence collected should be packaged separately. As a rule of thumb, each item of evidence should be packaged in a paper product, unless it is liquid, in which case it should be placed in a glass vial. Paper product means paper fold, paper bag or cardboard box. The remains of the deceased should be packaged separately, if that is the way they were found. If the body is intact, use a wooden backer board, white sheet and new body bag.

Once the surface of the site has been cleared of all remains and evidence, then recheck the area with a metal detector. Assuming there are no further indications from the metal detector, the next step is to examine and excavate the top 15 cm of soil for any further evidence or bones. In some instances the remains have gone through a self-burial. Objects placed on the surface of the ground may work their way into the ground. The extent to which this may occur will depend on the object's weight, ground density, terrain of the area, time elapsed and weather conditions.

The best method to use in removing the top several centimeters of soil is to cut the area into strips about 15 cm wide and remove the soil from the strips a section at a time. This material should then be sifted with a fine sifter. A sifter of 0.3–0.6 cm should be used so projectiles or teeth won't be missed. Once this is completed, the maximum amount of evidence and remains can be confidently obtained.

Contaminants can be contained in the soil beneath the remains; therefore, a sample of soil should be collected for further analysis by a forensic laboratory. A standard of the soil from a general area close to the remains must also be collected and submitted for analysis.

Finally, recover all other evidence or body parts in the area outside of the recovery site. This recovery should be handled and processed as thoroughly as any other outdoor crime scene.

The length of time from the initial search, to the location of the remains, to processing completion may be several days or weeks. Consequently, weather conditions must be considered. Plan accordingly!

Excavation Techniques

The basic procedures used in surface recoveries also apply to excavations; the difference is that the majority of the evidence and the remains are below ground level.

Once the burial site has been located and defined, the method of excavation needs to be chosen. Three methods of excavating the ground around the body, and ultimately the body itself, are recommended:

- *Hole* As the name indicates, a hole is dug, uncovering the remains as the soil is removed from over and around the body.
- *Trench* A trench is dug next to the remains to a depth of 0.5 m below the upper body level. The trench must be at least the length of the body and approximately 0.5 m wide. This trench will provide sufficient room to work and collect evidence and the remains. Using this method, three of the four walls of the grave can be defined.
- *Table* A table is dug by trenching all around the body, usually leaving a table approximately 1.25 m wide by 2 m long and extending 0.5 m beyond the depth of the body. This method will leave all four walls of the grave intact, as well as providing sufficient room to work around the body.

Because of the ease and comfort it provides while removing the remains and evidence, the table method is preferred. Regardless of which method is used, the position of the body under the ground must be estimated prior to excavation. This is not as difficult as it sounds. Based on the parts of the body which are visible, overestimate the position of the body and dig around it.

As with any of these methods, soil in strips approximately 30 cm wide and 15 cm in depth should be removed. The soil should be hand-checked and sifted as the different layers are removed. Having one qualified person in the pit and at least four other people using the sifters will maximize the amount of evidence recovered.

Anything that is not soil could be evidence or bones. Coins from victims' and suspects' pockets, a wine bottle cap that can be physically matched to a wine bottle found in the suspect's vehicle, skin tissue with ridge detail that can identify the ring finger of the victim, soy beans and corn stalks that provide a time element of the burial, magazines that can also provide a time element, and a whole host of other unusual items, not excluding weapons and projectiles, can be found. Any and all forms of evidence can be found in a grave site.

The easiest method of removing the body is to wrap it in a white sheet and place it into wooden backer board (all fire departments use them) before removing it from the grave. This will keep the body intact and make transportation easier. Once the body is removed, check the ground under it for possible footwear prints from the suspect in the soil; stranger things have happened! Several centimeters of the soil beneath the body must also be removed and sifted

again to locate evidence, bones, projectiles and teeth which may be present.

Often, buried bodies are literally yanked out of the ground and taken away from the scene with no thought to evidence, either in the grave or on the body. Just because a body is buried does not mean it cannot tell a story or point a finger at the murderer. Why make the job more difficult than it already is? If this was a fresh homicide scene and the body was in a parking lot, wouldn't everything at your disposal be used and everything possible to process the scene be done? Then why is it that, when a body is buried, investigators often have a different attitude? Probably because it is so unfamiliar. It is hoped that this manuscript will provide insight into the possibilities for physical evidence and the appropriate techniques to be used when recovering human remains which are buried and/or badly decomposed or skeletonized.

Take your time in the recovery of the remains and try to plan for the welfare of workers, the changing weather conditions, equipment needs, and 24 hour security at the scene.

Forensic Anthropology Recommendations

Because the remains in question are usually badly decomposed or skeletonized to some degree and little soft tissue may remain, an essential member of any recovery team is a forensic anthropologist. These forensic specialists usually have postgraduate qualifications in physical/biological anthropology, with education and experience in the recovery, identification and evaluation of remains suspected to be human. Their knowledge and experience can be critical in any case where decomposed and/or partially or completely skeletonized remains are expected. In cases involving badly burned remains, a forensic anthropologist *at the scene* can provide investigators with information that may be unobtainable once the remains are removed.

Forensic anthropologists' detailed knowledge of the human skeletal system allows them to provide information valuable in answering several important investigative questions when decomposed or burned human remains are involved:

1. Are the remains in question human?
2. Do the human remains represent a forensic case ($\leqslant$50 years since time of death)?
3. Who is the victim (i.e. age, gender, ethnic origin, stature, etc.)?
4. Do the remains exhibit any indications of pre- and/ or postmortem trauma that may assist the medical examiner or coroner in cause and/or manner of death determinations?

The ability of the forensic anthropologist to answer these questions with confidence, is dependent upon not only their experience level but also the amount of the skeleton and the specific skeletal elements (bones) recovered.

The majority of cases in which a badly decomposed body is discovered outside or suspected to be buried are likely to be homicides. As in all other crime scenes, this must be assumed. We cannot go back and treat the crime scene differently once we have left. Approximately 90% of all homicides are solved when the victim is positively identified. Therefore, the ability of the investigators to identify the victim will impact their ability to solve the case. However, the ability of the forensic anthropologist to assist in the positive identification of the victim is dependent upon the investigator's/crime scene personnel's ability to recover as much of the skeleton as possible.

In an adult human skeleton, recovery personnel, under ideal conditions, may find up to 206–211 individual bones. In infants and young children, many more individual bones should be expected. The bones of children have not all fused, as they have in adults. The shafts of the major long bones (bones of the extremities, e.g. arms and legs) are usually all present at birth. However, the ends of the bones (epiphyses) usually do not appear until sometime after birth and before puberty. As the child ages, not only do the shafts of the bone increase in length and diameter, but the epiphyses also increase in size. Eventually, the epiphyses fuse to the shaft to form an 'adult' bone. This occurs in most individuals by the age of 25 years.

In order to recover as much of the skeleton as possible, recovery personnel must be generally familiar with the normal components of the human skeleton. If a forensic anthropologist is not available for assistance in the recovery, investigators can get a general idea of what bones to expect from, for example, Bass's *Human Osteology: A Laboratory and Field Manual of the Human Skeleton*. Recovery personnel should use such a field manual to assist them in not only recognizing what bones are present, but, most importantly, also identifying whether key bones are missing.

While it is important to recover as much of the victim's skeleton as possible, recovering some parts of the skeleton will increase the forensic anthropologist's ability to assist in determining the victim's identity (i.e. age estimation, gender determination, stature estimate and ethnic origin if possible) and provide information useful to the medical examiner or coroner in determining cause and/or manner of death.

When a surface recovery is being conducted, a visual examination of the skeletal elements present

should be carried out prior to the removal of any remains. If it is determined that skeletal elements are missing, a more detailed, systematic search must be conducted. The extent of the search, as well as the type of search technique to be used, will be highly dependent upon the terrain and any observed indications of animal activity. Regardless of the search technique used, it must be systematic and not haphazard or random.

Owing to the presence of numerous, warm, moist orifices for flies to lay their eggs, the skull is typically the first part of the skeletal anatomy to decompose. As a result, the skull can often be found a distance from the main concentration of remains. Gravity, if the body is on a sloping hill or any incline, and the activity of scavenging animals, such as rodents or canines, can result in the skull being displaced after it has decomposed enough to be easily separated from the rest of the body. Similarly, if skeletal elements are determined to be missing from a body above or near a creek or stream, a search of the creek or stream bed will be required in the direction of water flow. Heavier bones, such as the long bones or pelvis, may be found closer to the original site; lighter bones, such as the bones of the hand, wrist, foot, ankle or the scapula and clavicle, may be found quite a distance downstream, depending upon the current. Forensic evidence may lie between the skull and main concentration of remains. Consequently, very detailed searches may need to be conducted in these areas.

Forensic odontologists are capable of positively identifying the deceased by comparing dental radiographs/records of the suspected victim to the teeth and bony structures surrounding the teeth that are recovered. The unique variations of one tooth may be all that is required to positively identify the deceased. It is therefore imperative that all teeth be recovered. It may be the one tooth that is not recovered that is the only tooth that can positively identify the victim.

A visual examination of the upper and lower jaw will indicate whether a tooth or teeth were missing before or after the death of the individual. The bony structures (sockets) that surround the teeth and help hold the teeth in place are reabsorbed if a tooth is lost premortem. The outer walls of the socket are not discernible. In teeth lost postmortem, the walls of the socket are identifiable and the 'socket' appears empty. Sometimes dirt can fill the socket. Use a soft instrument, like a toothpick, to gently push some of the dirt out. If the walls of the socket are identifiable, the tooth was lost postmortem. If is determined that 'missing' teeth were lost before death (i.e. no visible socket), there is no need to search any further for them. However, if teeth have been lost after death, they may be recovered from the scene.

For instance, if a skull is found away from the other parts of the body, it is likely that some teeth may have been dislodged/'lost' as the skull rolled away from the rest of the remains. Teeth can also be dislodged as animals carry away the skull. If open tooth sockets are observed, then a systematic search (usually with the searchers on their hands and knees) will need to be conducted in the area between the skull and main concentration of remains. More specifically, this area must first be divided into a grid using string. The debris in each grid should then be removed and sifted. Any tooth or other physical evidence found must be documented by photographs, measurements and sketches. Once documented, the evidence must then be placed in a paper bag, which is then properly marked and sealed to establish chain of custody. Once the surface debris has been sorted, 5–15 cm of soil from each part of the grid will need to be removed and sifted. Teeth that may have settled in the dirt may possibly be recovered in this fashion. The grid location of any tooth or piece of physical evidence recovered should be documented.

If the remains are buried in a shallow grave, it is possible that scavengers may have disturbed the burial and dragged individual bones or parts of the body away from the grave site. Any indication of soil disturbance, such as a small hole, should result in an expanded, systematic search of the area. As when surface remains are not complete, special attention should be given to identifying any possible paths used by animals. Once identified, these paths should be extensively and systematically searched. It is also possible that a scavenging animal, such as a coyote, fox or even a skunk, may carry parts of the body back to its den. Therefore, look for dens along the way. Any dens identified should be carefully investigated for the presence of remains.

When the deceased is an infant or young child, recovering as much of the skeletal elements as possible is critical in providing the forensic anthropologist with sufficient remains to make an accurate estimate of age. Owing to the underdeveloped nature of key characteristics, the gender and ethnic origin of prepubescent individuals cannot be determined with confidence. Emphasis should be placed on identifying and recovering the individual components of the skull, as well as, particularly, the long bones and epiphyses. Epiphyses are typically missed in recoveries because they resemble gray, porous clumps of dirt. Individuals working the sifters must pay special attention to recovering these elements. Although it is highly recommended that a forensic anthropologist participate in the actual recovery, if one is not available to assist in the recovery of an infant or child, collect all dirt and debris and submit as evidence to

the medical examiner, coroner or crime laboratory. Hopefully, a forensic anthropologist can examine this evidence in the laboratory. The specific location from which the dirt and debris was removed must be documented using standard procedures.

The forensic anthropologist may also be asked to assess the remains for evidence of trauma. It is important to identify and accurately document all indications of trauma. Documentation of premortem trauma may assist in the positive identification of the deceased. Identification of trauma which may have occurred around the time of death may be significant in the medical examiner's or coroner's determination of manner and cause of death.

Not all trauma will leave its trace on bone. Some trauma may only affect the soft tissues and not survive the process of decomposition; therefore, it is possible (but not probable), that in violent deaths no indication of trauma is discernible on the skeleton. However, several areas of the skeleton are critical if possible fatal injuries are to be identified and evaluated. Fatal injuries usually occur in the head and thorax (chest). Consequently, the skull, bones of the neck and upper back, shoulder and ribs should be handled very carefully. Any unintentional damage which occurs to the bones during recovery should be documented. Mistakes do happen, and can be intensified if they are not identified to the forensic anthropologist. Estimating when the fracture or injury occurred can be difficult, even without additional damage to the remains. Informing the medical examiner/coroner/crime laboratory of a mistake will save the forensic anthropologist valuable time.

Damage to other important skeletal elements can occur if recovery personnel do not handle the remains appropriately. The skull and mandible are important not only in gender determination, but the face, in particular, is also critical if the forensic anthropologist is to derive an opinion concerning ethnic origin. Often, measurements of the skull are taken to assess gender and ethnic origin. An incomplete or damaged skull will potentially eliminate the forensic anthropologist's ability to take accurate measurements.

Never pick up the skull by the orbits or cheek bones. These are very fragile areas where fracture from injury commonly occurs. Inadvertent damage caused to these areas may diminish the forensic anthropologist's ability to identify and appropriately evaluate fracture patterns. As previously discussed, any damage done to the remains during recovery and handling should be documented and communicated to the forensic anthropologist.

Inadvertent damage can also occur to the bones during transportation from the recovery site to the crime laboratory or medical examiner's/coroner's office. To avoid such damage, all long bones, the pelvis, the skull and the mandible should be packaged in separate bags. Never place any other bones in the same package with the skull, and never place other packages or equipment on top of the package that contains the skull.

An experienced and well-educated and trained forensic anthropologist is a valuable member of a recovery team. The amount of information that can be obtained from the recovered remains will depend upon the amount of the skeleton recovered, the condition of the remains and the specific remains recovered. The more that is recovered, the greater the amount of significant investigative information that can be obtained by these specialists.

Exhumations

An exhumation or disinterment is the removal of a legally buried casket from a cemetery. Persons legally buried may need to be exhumed because of possible misidentification or the recovery of new information pertinent to the cause of death. These conditions usually occur because the body was not properly processed before burial.

Exhumations are not complex. A court order or written consent must be obtained before any exhumation. Before beginning the exhumation, detailed information about the alleged deceased, name, age, height, weight, sex and cause of death, will be needed. A positive identification of the deceased will be required. The grave must be photographed prior to digging, during the exhumation and after the casket is removed.

A soil sample from the grave site will also be needed to prove that any toxins found in the body did not leach in from the surrounding ground. Soil samples should be taken from all four sides of the grave as well as from beneath the vault. A soil standard is also required and should be taken from the area adjacent to the grave.

The casket and body should be removed and transported to the facility where the autopsy will be performed. Once the autopsy is completed, the body and casket are returned to the burial site. The entire sequence is recorded photographically.

Additional Information

General descriptions of excavation and surface recovery procedures have been provided. The particular method used will depend on the terrain at the recovery site. The secondary area of the recovery site

should be checked for items of evidence, such as tools used to bury the body, clothing, tire tracks, footwear prints and trash left by the suspect. Prior to beginning excavation, contact utility companies for the location of any buried cables or pipes. If the weather dictates, have an enclosed tent available to cover the grave. If there is no shade, provide at least a canopy to cover the workers. Heat exhaustion is not an uncommon problem.

A variety of shovels can be used during excavation; however, as you get closer to the remains, change to smaller digging instruments, such as a hand trowel, bamboo stick or dental instruments. The diggers will work faster than the sifters. Therefore, a sufficient number of people should be available to sift the removed soil. Try to build sifters so they are 1.25 m above the ground for the comfort of those sifting. Use clean plastic buckets to remove the soil from the excavation and transport it to the sifters. Document where each bucket of soil came from in the grave site. The specific location of any evidence recovered can be documented in context. The sifters should be approximately 0.6–1.2 cm mesh. Educate the sifters to recognize evidence. Work only during daylight hours and give the workers sufficient break periods. Do not tire them out on the first day. Boredom in the sifters will cause evidence to be missed.

Water is needed at the recovery site for drinking, cleaning hands and equipment, and for cleaning items of evidence.

If the suspected burial site is on private property, a search warrant or written consent is needed.

The news media will try to get as close as they can to the burial site. Use police barrier tape to secure and identify the outer perimeter of the site. Watch for low-flying helicopters with reporters. Turbulence from helicopter blades can cause tents used for shielding workers to collapse or evidence to be blown away.

Security must be provided at the site on a 24 hour basis. Individuals will try to remove ‘souvenirs’ during the night. In doing so, they threaten the integrity of all remaining evidence. Security should be provided on site to prevent the theft of equipment.

Digging deep may result in a water table or a pocket of water being hit. A sump pump and a generator will be required to remove this water. Be prepared in advance to locate a sump pump, if needed.Recovery sites are all different and investigators must be able to adapt to them. Consequently, prior planning is critical. Equipment must be located in advance and must be available on weekends and holidays.

If you are unsure how to recover these types of remains properly, *don’t do it!* Anybody can dig a body up, but, without the appropriate education and training, few can find or identify the evidence.

The team concept cannot be overemphasized when recovering human remains that are buried and/or badly decomposed or skeletonized. In addition to the all-important first officer at the scene, the crime scene technician, the coroner/medical examiner, the forensic pathologist, the investigator, the forensic anthropologist and the state’s attorney (or prosecutor), there are a few other individuals whose expertise may be of great value to the case:

- *Forensic entomologist* A court qualified person who specializes in the study of insects and applies his or her knowledge and skills to legal cases. Knowledge of the life cycle of flies (eggs, maggots, pupa and adults) and beetles, potentially enables the time of death to be refined. A protocol for the collection of specimens will need to be followed; consult with the forensic entomologist.
- *Forensic odontologist* A court qualified dentist who applies knowledge of teeth and dental variation to legal matters. He or she is capable of providing a positive identification by comparing the deceased’s teeth with known dental records.
- *Forensic sculptorist* A court qualified person capable of performing a facial reconstruction from a skull.

Without these experts many cases would remain unsolved due to a lack of positive identification of the deceased, or a lack of a more accurate time of death. The time to locate these experts is now, not when the remains are found. Remember the old Boy Scout motto, ‘Be Prepared!’.

Finally, the investigator who can do everything him- or herself usually has several unsolved cases but a team of experts working together has few, if any, unsolved cases. The choice is yours.

In Summary

This text was written as an overview of recovery techniques and to provide investigators with insight into the physical evidence that is obtainable when recovering human remains that are buried and/or badly decomposed or skeletonized. Its sole purpose is to encourage the reader to seek additional education and training and refer to other publications on this subject. Always remember to take your time, record everything (notes, photographs and sketches) and do not assume anything!

See also: **Anthropology**: Overview. **Crime-scene Investigation and Examination**: Recording; Major Incident Scene Management. **Odontology**. **Anthropology**: Excavation and Retrieval of Forensic Remains.

Further Reading

Bass WM (1998) *Human Osteology: A Laboratory and Field Manual of the Human Skeleton*, 4th edn. Columbia, MO: Missouri Archeological Society.
Catts EP and Haskell NH (1990) *Entomology and Death: A Procedural Guide*. Clemson, SC: Joyce's Print Shop.
Fisher BAJ (2000) *Techniques of Crime Scene Investigation*, 6th edn. Boca Raton, FL: CRC Press.
Geberth VJ (1990) *Practical Homicide Investigation*, 3rd edn. New York: Elsevier.
Haglund WD and Sorg MH (1997) *Forensic Taphonomy: The Post-Mortem Fate of Human Remains*. New York: CRC Press.
Killam EW (1990) *The Detection of Human Remains*. Springfield, IL: Thomas.
Morse D, Duncan J and Stoutamire J (1983) *Handbook of Forensic Archaeology and Anthropology*. Tallahassee: Bill's Book Store.
Sansone SJ (1998) *Police Photography*, 4th edn. Cincinnati: Anderson.
Skinner M and Lazenby R (1983) *Found! Human Remains*. Burnaby, BC: Archaeology Press.
Spitz WV and Fisher RS (1993) *Medicolegal Investigation of Death*, 2nd edn. Springfield, IL: Thomas.

Scene Analysis and Reconstruction

L E Sharman, 12 Brook Street, Lower Hutt, New Zealand
D A Elliot, Institute of Environmental and Scientific Research Ltd, Auckland, New Zealand

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0460

Introduction

The authors work, or have recently worked, in a civilian laboratory that has always provided crime scene services for the police. The scene is secured and controlled by police, and the laboratory staff are on call to provide scientific expertise at the crime scene on the request of police. Typically, staff attend homicides, but they also attend any serious crime that, in the opinion of the officer in charge of the case, requires scientific input into reconstruction, testing at the scene, or evidence collection. This article addresses the broad issues of crime scene reconstruction from the point of view of a scientist, with an accent on the philosophy of forensic science investigation, types of personnel involved, areas of knowledge most applicable, and the process. The specifics of the scientific techniques applied are addressed elsewhere.

Nature of the Problem

The investigation of a putative crime scene addresses two questions:

1. What occurred at this scene, and was there a criminal offense committed?
2. Is there evidence of the identity of the offender(s) and other parties present?

The first question is addressed under the heading 'reconstruction'. It is always addressed, even if the focus of the investigation is on the second question. Without a reconstruction of the events that is close to the actual occurrences, evidence relating to the identity of an offender will not be gathered intelligently and much will be missed.

For the scientist at the scene there will be two sources of information: (1) the information from the police, which will be based on their understanding of statements of witnesses, complainants, suspects, and reports from doctors and pathologists; and (2) the marks, stains, traces, damage and items *in situ* at the scene. ('Scientist' is taken to mean a person who is usually a graduate or postgraduate in a physical science and who has a theoretical understanding as well as a practical base of the science he or she practices. A 'technician' has a practical expertise, but is not normally expected to develop theoretical aspects of his or her area of expertise.) The information from the police will be tentative, partial and developing throughout the period of the investigation.

As the investigation of a crime develops, there will be postulation of hypotheses by all investigators. Some will be checked and rejected immediately by cursory observation or an additional piece of information; others will stand for some time, until there has been gathering of data or completion of laboratory analyses, before being rejected. Some requests and suggestions by the police will be for tests that are impossible or extremely unlikely to be successful with the present technology: 'If he was here we might find fibers from his clothes in the vacuum sweepings?' But there will be a few hypotheses remaining that are established as 'not disproven'. The ideal to strive for is to eliminate all but one conceivable sequence of events as being consistent with all the material evidence, observations and evidence from witnesses. This, of course, is not always achieved, but many 'wild-goose chases' caused by wrong or misinterpreted information can be stopped by applying science to a reconstruction.

Some legal experts have questioned sampling with a hypothesis in mind, on the basis that the scientist or scene examiner has prejudged a situation and the subsequent forensic science merely functions to confirm (police) investigators' theories: 'find a suspect, then collect evidence to convict him'. However, a hypothesis–test cycle can never prove a hypothesis, merely leave it 'not disproven'.

Most hypotheses enunciated at a crime scene never reach anywhere near a courtroom – as indicated above, they are eliminated during the investigation. Hypothesis-free sample collection, or for that matter exhibit testing, is impossible. It would imply that all items at a scene and every square centimeter of wall, floor and ground at a scene have had all possible tests applied. Such a plethora of information, even if it could be collected, would be meaningless. Contrariwise, collection of any item for any test, presupposes a hypothesis about the role of that item in a crime event. To canvass all possibilities adequately, and not be misled by initial assumptions, it will be most important to be open minded at all stages of an investigation. The investigator must not commit the recording and collection of evidence too early to a supposition of the investigators or place too much credence in the statements of witnesses.

Who Should Carry Out a Crime Scene Reconstruction?

In the authors' jurisdiction, a forensic scientist is nearly always called to the scene when crimes taken most seriously by the community are discovered. These include homicides and some rapes. The ideal is to have scientists always attend all scenes of crime, but this is probably not cost-effective.

The scientists attending the crime scene will have been trained in crime scene examination. Trainees may attend scenes as part of their training, but they will take the role of assistants. The scientific input into a crime scene reconstruction will be done by the scientist, working closely with police investigators and police technicians and experts.

Why Should a Scientist be Called?

It is at the crime scene, before significant disturbance has occurred, that informed hypotheses as to the events are formed. These in turn will guide the sampling. If forensic science is indeed a science, a scientist must attend all major or complex crime scenes. In no other field of science would the professional scientists contemplate allowing the framing of hypotheses and the questions to be addressed to be left to nonscientists. In most scientific research a senior researcher frames hypotheses and designs an experimental program; the actual execution of the program is carried out by graduate students and technicians.

In contrast, in much of the practice of forensic science worldwide, the operations of the professional forensic scientist are often limited to such work as can be done on portable items submitted to the laboratory. The truly scientific work – framing of questions to be asked, identifying the problems to be addressed – is left to the least scientifically trained section of the forensic community: crime scene technicians or detectives. It says much for the ability and persistence of those officers that forensic science works at all.

A trained scene of crime officer (SOCO) can cover many fields of expertise – fingerprint location and collection, photography of crime scenes, as well as collection and preservation of forensic samples – and as such is very valuable at crime scenes. Scientists, on the other hand, have a depth of knowledge and an understanding of the principles of their science, but tend to be specialized in their fields.

Experience is a major factor in the effectiveness of an expert, but experience alone is insufficient to make that person an expert. An expert is capable of recognizing exceptional circumstances, framing hypothetical reconstructions and eliminating or accepting them on the basis of theoretical knowledge coupled with tests applied at the scene or back at the laboratory. For example, in fire investigation, well-formed opinions must be based on an understanding of the chemistry, physics and engineering aspects of fire spread. Having been a senior fireman does not make that person an expert fire investigator. That person's experience could however be an excellent base on which to build theoretical and conceptual training in order that he or she becomes an expert.

Science of Reconstruction

The crime scene is completely unpredictable; the original crime can involve any number of activities and instruments. The scientists or investigators attempt to reconstruct those activities from their evidential traces by applying a wide range of sciences, including a few peculiar to forensic science itself. These may include some or all of the following:

- forensic pathology;
- bloodstain pattern interpretation;
- footwear print development;
- fingerprint enhancement and development;
- forensic 'ballistics' (inference of projectile trajectory and the firer's position from projectile traces);

- identification of firearm weapons from cartridge cases and projectile remains;
- photography.

Crime scenes involving fires should be examined by an expert fire investigator, assisted by fire and electrical engineers should the investigator lack those skills. Engineers may be involved with offenses involving machinery and vehicle accidents or homicides. The investigators may also need to call in botanists, entomologists, geologists, and so on.

The areas of expertise described above represent a range of skills, not the number of persons to be called. A forensic laboratory that supplies scene of crime services will train its scientists and technicians so that each person attending has a range of skills that minimizes the number of experts attending a crime scene. The laboratory will also have working relationships with experts, as listed above, who may be required on an occasional basis.

When a number of experts are working at a crime scene, they are to be coordinated so that (1) they are aware of each others findings and theories, and (2) each expert carries out his or her examination with minimum disturbance of trace evidence of interest to others. The several experts involved should briefly confer so that the best order of examination can be agreed, and precautions to protect potentially fragile evidence can be established. That conference, which may only take a few minutes, could be called and coordinated by the police officer in charge of the scene.

It should go without saying that the decision as to priority for examination and evidence collection should be based on the logic of the type of evidence expected and its fragility, not the relative rank, seniority or status of the various experts. For example, a frequent practical problem is the establishment of priority of fingerprint evidence and blood evidence that may give a DNA profile. Even when the two types of evidence occur in proximity, with good will and cooperation both experts can usually obtain optimum samples.

The Process

The initial call

A laboratory offering crime scene examination as a service must make staff available on 24 h call. The staff will typically be available on a roster, and will have a mobile telephone and/or a pager. When a scientist is called to a scene, there should be discussion between the expert and a police officer before the scientist ventures out. Can the scientist do anything immediately? Is lighting available? Has the staff member the appropriate expertise? Will the scientist have access to the scene on arrival, or will there be a protracted wait while photographers or fingerprint examiners complete their work? What are the circumstances (as understood at the moment)? An account of the circumstances, as understood at the time, may indicate the type of expertise that will be required, but the scientist should be ready to alter this initial opinion as the circumstances clarify themselves through the investigation.

Equipment

Scientists who attend crime scenes will have their own kit of equipment, either personally maintained or maintained and monitored by the laboratory. It will include packaging material, gloves, protective clothing and footwear, a torch, a lens, measuring tape, and spot-test reagents known to be currently active. The laboratory will probably also maintain specialized kits, such as spray kits for luminol. A van or similar vehicle set-up for scene examinations, containing maintained general kits, is useful, although our experience has been that well-boxed kits that can be carried into a small van or station wagon allows more flexibility than a dedicated van.

Initial inspection

The scene investigator should be prepared to do nothing but look and listen for a considerable time after arrival. The various police accounts of the events will be reiterated several times. These accounts may be changing very quickly at the initial stages. The reconstruction at this stage is only tentative and the scientist's investigation may greatly modify it. The only reason to carry out any action immediately is to save or preserve an item of potential evidence that is endangered by deterioration if left in its environment. Evidence may be damaged by a number of causes, which include water, heat, household pets and, particularly, others at the scene.

Footwear prints are particularly liable to be lost or obscured, and if there is a chance of footwear print evidence being present, immediate steps must be taken to protect that evidence if protection is not already in place. Police these days have an appreciation of the usefulness and fragility of shoeprint evidence. Therefore, those initially attending a scene can be expected to secure it and clear an access path to the body and the immediate area of interest with minimum disturbance of the potential evidence. Overshoes that do not add further shoe prints are to be worn by the few personnel that must be present at the

scene, and observation and recovery of footwear prints should often take precedence over even fingerprint recovery. Prints of the footwear from those personnel known to have been at the scene are to be routinely collected and retained.

Other evidence, such as fingerprints, bloodstains, tissue remains and dust marks, are also fragile or may be lost by careless investigation.

This initial inspection, becoming aware of the scene itself, and how items at the scene may have been ordered prior to the crime, is a critical first step. Major features will have been noted, and an overall strategy for observing, recording and collecting evidence will be formed. There will be questions to address to the investigators initially at the scene or who have interviewed persons involved. Now is the time to confer with others carrying out other specialized functions at the scene: fingerprint technicians, pathologists, photographers, the officer in charge of the scene, or surveyors. Secure an agreed process for examination of the scene and collection of the items. Are you satisfied that the cleared pathways are really free of trace evidence? Should staff wear footwear that will protect the scene from further loss of footwear print evidence?

Now is the time to begin reconstructing the events.

Bloodstain pattern interpretation

This is a specialty of forensic science, and is treated elsewhere. From theoretical consideration of the mechanics of a liquid striking a surface and empirical study of patterns arising from dropped and splashed blood, it is possible to infer many facts that will assist reconstruction of events and collection of the most relevant samples at a crime scene. These could include:

- type of weapon used, including approximate dimensions;
- number of blows;
- positions and movement of assailant and victim;
- evidence of movement of furniture in the room during and after the events in question;
- number of persons bleeding;
- when there is no body, whether the victim was alive or dead when he or she was removed from the scene;
- amount of blood lost: calculation of an approximate value for this can indicate whether or not a person may have died at a scene.

The person carrying out this type of examination must be trained and have proven skill in this area.

There are a number of standard works that address bloodstain pattern interpretation in detail.

Luminol

A supplement to blood splash interpretation is the use of luminol. Luminol is a reagent that detects substances with peroxidase activity, one of which is blood. The test is spectacular to the dark-adapted eye: the positive result appears as a faint blue-white luminescence (*not* fluorescence or phosphorescence, as often described) in the presence of blood. The reagent is sprayed over the scene when it is in complete darkness. The glow can be photographed with suitable fast film (ASA400 or greater). The locality of the objects in the room in relation to the luminol can be achieved by double exposure of the luminol, timed exposure and a flash photograph. Marking the visible part of a stain with brass drawing pins can also assist. Metals and some paints can interfere by causing the reagent to 'flash' in an instant reaction and hence obscure the faint but persistent glow of the positive reaction.

Obviously luminol is not an absolutely specific test for blood in itself. However, while things other than blood may in principle show peroxidase activity, the activity in blood is unique in its strength and persistence. Some freshly broken plants or squashed insects, for example, may show this activity, but the presence of these should be obvious to the careful and trained investigator. The interference by metals and paints is also quite different in appearance to an experienced investigator.

Thus luminol is invaluable to the investigator when all other evidence has been noted and collected. The test must be the final action at a scene because the action of spraying the water-based reagent over the scene will no doubt damage or destroy other trace evidence. What luminol does, when used cautiously, is to illustrate the overall distribution of blood, make the investigator aware of blood that has been mostly removed, and draw attention to hitherto unsuspected prints and patterns. Even when scenes are examined carefully, once attention has been focused by a luminol reaction, small traces of visible blood can often be found for subsequent testing for confirmation of blood.

Footwear prints

The fashion of wearing rubber-soled trainers or similar footwear has been a boon to the crime investigator. This is because of the clear prints the soles of these shoes leave, the wide variety of type of sole, and the tendency of the sole to be easily damaged, which makes each sole unique. Most footwear print evidence addresses the problem of 'who did it?' rather that 'what happened?', but immediate and valuable information is available by identifying how many

different types of footwear prints are present (after allowance is made for prints from witnesses, medical and ambulance personnel, etc.). The prints can be in a wide variety of forms and substrates and they can be enhanced by chemical and physical methods.

If there has been an assault involving loss of blood, shoes inevitably print this blood, often invisibly, but visualized by luminol or other enhancement methods. The choice will depend on whether it is the detail of the print that is required or merely the following of a trail. This visualization can track pathways through a building and may indicate a departure trail.

Weapon and toolmarks

These may be dents, paint smears or soil smudges. They may or may not be associated with blood. Think carefully how and where a weapon or tool may have been used and remember to look for marks caused by the back-swing or leverage. *Never* use the suspected weapon or tool at the crime scene in a reconstruction experiment during the investigation.

Points of entry

This is an obvious part of reconstruction. If a building has been entered through a forced opening, location and reconstruction of the exact mode of entry can assist location of trace evidence, footwear prints and toolmarks. Collect samples from paintwork contacted and broken glass, for later comparison with traces on a possible offender.

What may have been transferred to the offender?

Traces from the point of entry have already been mentioned. A knowledge of the way blood has been splashed is useful here.

Experimental reconstruction

It is always illuminating to attempt to reproduce a hypothesized reconstruction back at the laboratory or at the crime scene itself. Experimental reconstruction at the crime scene is very informative but should be done only at a late stage of the investigation, when investigators are satisfied the scene has been cleared. Weapons and tools must be as exact copies of the supposed actual items as is possible. Actual exhibits are to be used in experimental reconstructions only when there is no alternative, and only if there is certainty that no further contact evidence remains on them.

Training of the Crime Scene Examiner

If the forensic laboratory is offering scene analysis as a service, examination of crime scenes should be a regular feature of the work. If a scientist is called to crime scenes on occasions when other investigators encounter and acknowledge problems with the reconstruction of a crime scene, the scientist may well have little to offer, simply because he or she has had too little opportunity and training in this field.

Every opportunity should be taken for trainee scientists who are going to become scene examiners to visit crime scenes with a mentor. Specific training and competence testing should be offered in the crime scene forensic science skills listed above. The laboratory training program should specify a number of crime scenes to be attended in the trainee role, with some specification of variety of type of scene, and there should be opportunity for the trainee to carry out work and be assessed by a supervisor/trainer.

As indicated above, the authors believe that crime scenes that require scientific analysis or reconstruction should be attended by graduate scientists, and thus at least some of the trainees at a laboratory should be graduates. However, there will always be a role for the crime scene technician, and realistically that technician will be doing at least some reconstruction, if only because it is difficult to conceive of proper selection of evidence material without implicit reconstruction.

An Illustrative Case

The following is a recent case involving one of the authors (DAE) that illustrates many of the points above:

In October 1996, a scientist from the Institute of Environmental and Scientific Research attended the scene of a homicide in a park in central Auckland, having been called to it by the officer in charge. As rostered scientist, and one with the expertise required, he became supervising scientist of this case. Also present was a pathologist. The victim was a female; her clothes had been disturbed, with her pants and tights pulled down, and her upper garments pulled up. One large rock lay on her left shoulder; another large rock lay by her arm. These were later determined to be 33 kg and 11 kg. Both rocks were bloodstained. The deceased's face was heavily bloodstained and injured. A condom was hanging from her mouth.

It was decided that further scene work was required before the body was removed. The police photographer was present during this and subsequent work. A scientist and a trainee scientist with expertise in shoeprint work attended the scene and made casts of disturbed areas around the body. The body was removed. The case scientist attended the initial part of the postmortem examination to discuss injuries and to advise on the collection of trace evidence and the removal of clothing. Interpretation of the blood

splashing and smearing on the body was also performed.

Upon returning to the scene, the area immediately around the body was searched and cleared by the scientist with the assistance of the police scene officers. The police grid-searched the larger area.

Combining the blood splash evidence on the body with that revealed at the scene, together with information from the postmortem examination, the scientist was able to reconstruct the events. It appeared that the victim was hit over the head more than once with one of the rocks, moved and hit again with one of the rocks. This was determined from the pooling of blood in the general area of the body and the sequential nature of the blood spatter. Semen was detected on the ground related to the final position of the body.

Luminol was used that night to trace the movement of the assailant. The scientist was assisted by police officers and a trainee scientist. The police arranged for street lights to be doused during this examination. A trail of 270 paces was established, taking the assailant from the body, through the park, and across a main road. The luminol trail indicated that the assailant had crossed the road by vaulting over railings; a bloodied palm print was detected on the railings. The blood was later profiled as the victim's; the palm print was that of the person eventually found guilty.

Appropriate scene work involving scientific and medical expertise and police skills used here helped reconstruct the events of that night, both the attack itself and the exit from the scene of the offender. Where corroboration was forthcoming, the reconstruction proved accurate.

See also: **Accident Investigation**: Motor Vehicle; Rail; Determination of Cause: Overview. **Basic Principles of Forensic Science**. **Serology**: Blood Identification; Bloodstain Pattern Analysis. **Causes of Death**: Overview; Scene of Death. **Crime-scene Investigation and Examination**: Recording; Collection and Chain of Evidence; Recovery of Human Remains; Packaging; Preservation; Contamination; Fingerprints; Suspicious Deaths; Major Incident Scene Management; Criminal Analysis. **Criminalistics**. **Fingerprints (Dactyloscopy)**: Visualization; Identification and Classification. **Fire Investigation**: Types of Fire; Evidence Recovery at the Fire-scene; Fire-scene. **Firearms**: Weapons, Ammunitions and Penetration; Range. **Engineering**. **Modus Operandi**. **Pathology**: Overview; Victim Recovery; Preservation of Evidence. **Pattern Evidence**: Footwear Marks; Bare Footprint Marks; Vehicle Tire Marks and Tire Track Measurement.

Further Reading

Bevel T and Gardner RM (1997) *Bloodstain Pattern Analysis: With an Introduction to Crime Scene Reconstruction*. Boca Raton, FL: CRC Press.

Bodziak WJ (1990) *Footwear Impression Evidence*. New York: Elsevier.

DeHaan JD (1997) *Kirk's Fire Investigation*, 4th edn. Upper Saddle River, NJ: Brady.

Eckert WG and James SH (1989) *Interpretation of Bloodstain Evidence at Crime Scenes*. New York: Elsevier.

Fisher BAJ (1993) *Techniques of Crime Scene Investigation*, 5th edn. Boca Raton, FL: CRC Press.

Fox RH and Cunningham CL (1992) *Crime Scene Search and Physical Evidence Handbook*. Boulder, CO: Paladin Press.

Freckelton IR and Selby H (1993) *Expert Evidence*. North Ryde, NSW: Law Book Co.

Geberth V (1983) *Practical Homicide Investigation: Tactics, Procedures, and Forensic Techniques*. Amsterdam: Elsevier.

Given BW, Nehrich RB and Shields JC (1977) *Obtaining Motor Vehicle Evidence from Tire Tracks and Tread Marks: A Complete Reference for Collecting, Recording and Analyzing Track and Tread Evidence*. Houston, TXL: Gulf.

Goddard KW (1977) *Crime Scene Investigation*. Reston, VA: Reston.

Heard BJ (1997) *Handbook of Firearms and Ballistics: Examining and Interpreting Forensic Evidence*. Chichester: Wiley.

Lee HC (1994) *Crime Scene Investigation*. Taoyuan, Taiwan: Central Police University Press.

Macdonell HL (1996) *Crime Scene Evidence: Blood Spatters and Smears and Other Physical Evidence*. Hamden, CT: Quinnipiac Health Law.

Saferstein R (ed.) (1982) *Forensic Science Handbook*, vols 1–3. Englewood Cliffs, NJ: Prentice-Hall.

Tilley N and Ford A (1996) *Forensic Science and Crime Investigation*. London: Home Office Police Research Group.

Vanezis P and Busuttil A (1996) *Suspicious Death Scene Investigation*. London: Arnold.

Suspicious Deaths

J Horswell, Australian Federal Police, Canberra, Australia

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0457

Cause of Death

To understand suspicious death the crime scene investigator should understand the various causes of death and the manner of death. The cause of death

refers to the disease or injury responsible for initiating the chain of events that results in death. Some examples of causes of death include cancer, heart disease, pulmonary embolism, gunshot wound(s), stabbing wound(s) and asphyxiation by strangulation or hanging. They may be acute or chronic. The cause of death should reflect the first event that sets into motion a number of sequential events, such as broken bones from a blunt injury leading to immobilization resulting in pulmonary embolism. In this case the blunt trauma is the original cause, whereas the pulmonary embolism is the final cause of death. One would expect the cause of death to read 'Cause of death = Pulmonary embolism due to blunt trauma injury to the skeletal system [here nominate the bones broken]'.

Manner of Death

The manner of death relates to the circumstances in which the death occurred, and is generally divided into five categories:

- *Natural* Death is due to a natural process or disease.
- *Accident* Death is due to the unintentional or inadvertent actions of the deceased or another person.
- *Suicide* Death is due to an intentional act or acts of the deceased who anticipates his or her resultant death.
- *Homicide* Death is due to the direct action of another.
- *Undetermined or unknown* The circumstances surrounding the death cannot be determined with reasonable certainty.

It is important for the crime scene investigator to be knowledgeable about death scenes. It is also important for a forensic pathologist to attend the scene, however this is not always possible, particularly in rural or remote communities. A heavy burden is therefore placed on the crime scene investigator who must have the knowledge and skills to understand fully the various causes and manner of death. This can only be achieved by visiting as many death scenes as possible. Scenes provide important information that may indicate cause of and means of death. The environment at the scene may help explain the condition of the body when it was discovered. If the deceased died while lying close to a heater, the body will decompose faster than usual. If the deceased collapsed on to an irregular or rough surface, this may explain minor injuries that are found on the body. Postmortem injuries should not be confused with antemortem injuries. Cause of death may be obvious – for example, a gunshot wound to the head – however, to determine the manner of death, further information will be required, such as whether this was an accident, a suicide or a homicide.

Death scenes should be treated as homicide scenes and dealt with as such until the contrary is proved. This can only be achieved through detailed examination and examination.

There are now two distinct tasks ahead of the crime scene investigator. The first is the technical recording and retrieval of potential evidence from the scene. Just as important is the second, the reconstruction in the mind of the crime scene investigator of the events surrounding the death. This article will discuss the technical issues first, and the reconstruction issues second.

Scene of Death

When a violent or suspicious death is discovered, the death scene should be secured immediately so that no one will have an opportunity to change it in any way. Indoor scenes may be easy to secure and protect. Outdoor scenes can be difficult. The more urban the scene, the more difficult it is to secure, as there will be a need for several scene guards and the setting up of cordons. The more remote the scene, the easier it is to secure. The weather and movement of predatory animals through common approach paths, and the maintenance of a scene log to record visits to the scene, add a dimension to the processing of a crime scene and the condition in which the deceased may be found. Outdoor scenes may be cooler than indoor scenes and the weather can have an impact as can the movement through a scene by animals. Any disturbances must be noted.

The crime scene investigator should take photographs and video recordings immediately before anything is moved. 'Stepping plates' may be used to secure a route to the body and preserve evidence. Notes and a sketch should also be made at this time. The deceased's location relative to other objects and structures within the scene is very important. The actual position of the deceased is plotted: the head and crutch are good points on the body to use for plotting its position. Accurate measurements and relative proportions must be noted to place items accurately within the scene in sketches made at the scene. Great care should also be taken to preserve any footwear marks or other significant evidence types around the body.

The deceased is the most valuable piece of potential evidence at any death scene. Hence, a systematic and thorough examination of the body should be under-

taken at the scene. Blood, weather conditions, location and poor lighting may mask faint injuries and trace evidence on the body; therefore the crime scene investigator should document in writing, by sketch and by photographic recording all information about the body that can be gathered at the scene. The remainder will have to wait until the postmortem examination. The attendance of other experts at the scene may also be considered, e.g. biologist, entomologist, etc.

The environment in which the body was found will affect the rate of body cooling. The wind conditions, temperature and the presence of any rain should be noted. The crime scene investigator will need to develop a general description of the deceased, including, where possible, the gender, race, age, height and weight.

One of the first most important questions that needs answering is: Did death occur at this locus? The position in which the deceased was discovered is of particular importance, as it will provide an indication as to whether or not the body was moved before being discovered. The presence or absence of rigor mortis or stiffness of the body, absent, minimal, moderate, advanced or complete, will help the crime scene investigator and the pathologist determine if the person died at that locus, in the position, as found. The crime scene investigator should inspect the body for postmortem lividity, hypostasis, or livor mortis. A pink-purple discoloration is usually present at the lowest point of the body. This is due to the settling of the blood by gravitation. Its location and state of fixation should be noted and photographed. For example, unfixed livor mortis blanches white when moderate pressure is applied, as opposed to fixed livor mortis, which remains the same color when pressure is applied. If livor mortis is noted on the deceased in areas not consistent with it forming in the lowest parts of the body, the crime scene investigator should consider the possibility that the deceased was moved after death.

The blood flow patterns should match the position of the body. If the scene is one of apparent violence, the blood flow patterns can indicate the use of a weapon.

Is there trace evidence at the scene consistent with the death having occurred at this locus? Does the body contain any trace evidence that is unusual for this locus; for example, mud on soles of shoes and/or grass and/or seed material found on the clothing when the body has been located inside a house? Is the death one that can be attributed to natural causes? Is there any external sign of violence? Is there anything amiss or out of the ordinary with the scene? Is there anything about the scene that arouses the crime scene investigator's suspicions?

The crime scene investigator should consider several hypotheses and then set out to prove or disprove each one. The physical evidence present, togther with the known facts, should be sufficient to enable the crime scene investigator to develop a reasonable hypothesis as to what has happened at the scene; however, this is not always possible. Suspicions may not be aroused until the postmortem examination reveals something that was not apparent at the scene. The pathologist may, however, provide the investigator with a definitive indication as to the cause of death, giving the investigators leads to start their lines of inquiry.

Before the deceased is moved from the scene, the crime scene investigator should begin the examination of the body systematically, noting and photographing trauma and locating and removing potential trace evidence that may be lost on moving the body. The taking of body swabs and tape lifts may be considered at this stage. The deceased's hands, feet and head should be bagged using paper bags, as the use of plastic and any subsequent refrigeration will cause plastic bags to sweat. The bags should be large enough to be taped securely around the wrist, ankle or neck and allow 'ballooning' over the area to be protected. The body should then be placed into a new, clean body bag. If face down, it should not be turned over in this process in order to avoid 'flooding' wounds and thus damaging evidence. A thorough examination of the area that was previously obscured by the body may now be conducted. The deceased should then be transported to the mortuary for a full postmortem examination.

Postmortem Examination

The postmortem examination is normally conducted by an experienced forensic pathologist. The crime scene investigator should be present, as will the investigating officer or his or her delegate.

The deceased should be taken from the body bag and all packaging items should be retained for a subsequent search for trace evidence in the forensic science laboratory. The crime scene investigator should then record photographically the clothing and the condition of the body, and take a facial photograph for use in identification processes. The forensic pathologist should be taking notes and making diagrams.

The clothing should be inspected and any obvious trace evidence removed and packaged. Each item of clothing should be removed and packaged separately. It is at this time that wallets and property should be removed from the deceased's pockets and the con-

tents noted and the items packaged. It may well be that these are of no forensic value, as such, and thus may be released into the custody of the investigating officer. All items of clothing should remain in the custody of the crime scene investigator.

An external examination should take place next. Questions may already exist that need answers. Are there any marks of violence, such as stab or gunshot wounds, ligature marks, or restraints, injection sites, or blunt injury wounds? Are these injuries incompatible with life, or could the person have lived and moved for a time after sustaining the injuries, thus explaining some of the evidence located at the scene? Could the deceased have inflicted the injuries? Were there other persons involved? Do the injuries appear to have been made antemortem, at the time of death or postmortem? Bleeding into the tissues (bruising) usually means that the deceased was alive at the time of the injury. Are there any injuries that could be regarded as 'defense' wounds? These would indicate that the deceased was alive, alert and aware that he or she was being assaulted and aware of the life-threatening nature of the assault.

The crime scene investigator should search the body using oblique lighting. This is a method of highlighting minute particulate matter and/or bruising that, under normal lighting, is difficult to see.

At this stage of proceedings, the forensic pathologist should take fingernail scrapings, with each digit's scrapings being placed into a separate labeled container. The next samples to be obtained are hair samples, which should be recovered from a site of interest and packaged. These samples should remain in the custody of the crime scene investigator. A complete series of radiographs should then be taken. These may reveal broken bones, projectiles, foreign bodies, prostheses that have been inserted into the person while alive, and major deformities.

Identification may be an issue. Identification is important in both civil and criminal law. It provides proof of the person's death. It allows for:

- notification of next-of-kin;
- eventual disposition of the remains;
- completion of the death certificate;
- settlement of the deceased's estate;
- identification in criminal proceedings.

Visual identification, although the most frequently used method of identification, is subjective and can be unreliable as many people bear resemblances to each other. The person making the identification can also be so shocked by seeing the deceased that he or she can make a false identification owing to being in a state of denial. Scientific identification techniques may also be employed, including:

- A general description of the deceased: color of hair and eyes, height, the presence of scars, marks and tattoos.
- A general description of the clothing: type and color of each item.
- A forensic odontologist should attend the mortuary before the postmortem examination begins to chart and cast the teeth and to take what radiographs may be necessary for later comparison purposes.
- DNA samples (see below).
- Contents of the deceased's wallet.
- A medical records search for any major operation, broken bones or the addition of any prosthesis fitted during life, if a name can be suggested for the deceased.
- The last issue to be dealt with before the deceased is released is that he or she be fingerprinted. A fingerprint technician should attend the mortuary, take inked impressions from the deceased's hands, classify them and search for a match through their fingerprint files.

The pathologist continues with the postmortem examination, during which items of interest are recorded by the pathologist in his notes and diagrams, and photographically by the crime scene investigator, who should use a scale rule with any close-up photography. It is also at this stage of the postmortem examination that samples should be taken for blood grouping and DNA profiling, alcohol determination and drug and poison screens. Samples should also be taken for histopathology and microbiology if the forensic pathologist thinks that these are necessary.

Violent crimes typically result in a great deal of physical evidence being deposited on the deceased and the suspect. A thorough search should be made of the deceased or living suspect for any evidence, which may be expected to be extremely transient and easily destroyed or contaminated. Depending on the type of crime, for example shooting, stabbing, rape/murder or blunt injury beating, different forms of evidence may be present. Additionally, the specific area on the body in which the evidence is found is of extreme importance; for example, dried semen in the pubic hair of a deceased female, or firearm discharge residue on around the entry point to the chest is indicative of a contact wound. Consideration may also be given to other experts attending the postmortem examination (e.g. firearms expert).

In crimes of violence the crime scene investigator should insist on the taking of blood and other biological samples for grouping and DNA analysis; and blood, vitreous humor and urine for alcohol levels.

Summary

After all the information is to hand from the crime scene and the postmortem examination, those involved should be able to work out the cause and manner of death from the facts available. Although modern forensic investigation is advanced, there will be times when the crime scene does not provide information and the postmortem examination does not reveal a definitive cause of death. These are the difficult cases.

See also: **Causes of Death**: Overview; Scene of Death; Postmortem Changes; Sudden Natural Death. **Crime-scene Investigation and Examination**: Suspicious Deaths. **Identification/Individualization**: Overview and Meaning of ID.

Further Reading

Fisher BAJ (1993) *Techniques of Crime Scene Investigation*, 5th edn. Boca Raton, FL: CRC Press.
Geberth VJ (1983) *Practical Homicide Investigation*. New York: Elsevier.
Geberth VJ (1996) *Practical Homicide Investigation*. New York: CRC Press.
Goddard KW (1977) *Crime Scene Investigation*. Reston, VA: Reston.
Horswell J (1993) Crime scene investigation. In: Freckelton IR and Selby H (eds) *Expert Evidence*. North Ryde, NSW: Law Book Co.
Kirk P (1974) *Crime Investigation*, 2nd edn. Chichester: Wiley.
Snyder L (1973) *Homicide Investigation*, 2nd edn. Springfield, IL: Thomas.
Spitz WU and Fisher RS (1980) *Medical Legal Investigation of Death*, 2nd edn. Springfield, IL: Thomas.
Svensson A and Wendel O (1974) *Techniques of Crime Scene Investigation*, 2nd edn. New York: Elsevier.
Tuthill H (1994) *Individualisation: Principles and Procedures in Criminalistics*. Salem, OR: Lightening Powder Co.
White P (ed) (1998) *Crime Scene to Court – The Essentials of Forensic Science*. Cambridge: Royal Society of Chemistry.

Criminal Investigation *see* **Modus Operandi**.

CRIMINAL PROFILING

B E Turvey, Knowledge Solutions LLC, Watsonville, CA, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0754

Introduction

Criminal profiling is founded on a diverse history grounded in the study of criminal behavior, the study of mental illness, and the forensic sciences. It involves the inference of offender characteristics for investigative purposes, as well as for those that service the needs of the court. Criminal profiling methods tend to involve the use of either inductive or deductive reasoning, which has added to the confusion about the nature of criminal profiling in the popular media, as well as in the law enforcement community and the judicial system.

Profiling: A Brief History

Origins

The process of inferring distinctive personality characteristics of individuals responsible for committing criminal acts has commonly been referred to as ‘criminal profiling’. It has also been referred to, among other less common terms, as behavioral profiling, crime scene profiling, criminal personality profiling, offender profiling and psychological profiling. There is currently a general lack of uniformity or agreement in the application of these terms to any one profiling method. As a result of this, and the varied professions involved in the profiling process, these terms are used inconsistently and interchangeably. For our purposes, we will use the term criminal profiling.

A common aspect of the criminal profiling process is understanding and classifying criminal behavior,

and by extension criminals themselves. The renowned Italian physician, Cesare Lombroso, is generally thought to have been of one of the first criminologists to attempt to classify criminals for statistical comparison. In 1876, Lombroso published his book *The Criminal Man*. By comparing information about similar offenders, such as race, age, sex, physical characteristics, education and geographic region, Lombroso reasoned that the origins and motivations of criminal behavior could be better understood.

Lombroso studied 383 Italian criminals. His evolutionary and anthropologic theories about the origins of criminal behavior suggested that, based on his research, there were three major types of criminals:

- *Born criminals*: degenerate, primitive, lower evolutionary reversions in terms of physical characteristics.
- *Insane criminals*: those who suffered from mental and/or physical illnesses and deficiencies.
- *Criminaloids*: a large general class of offenders without specific characteristics, not afflicted by recognizable mental defects, but whose mental and emotional make-up predisposed them to criminal behavior under certain circumstances; a concept which has been compared to the diagnosis of psychopathic personality that came later from the psychiatric community.

The suggestion of Lombroso's work was that certain types of criminals could be visually identified by investigators, merely by recognizing their criminal features. According to his theories, there are 18 physical characteristics that are indicative of a 'born criminal', when a person has at least five or more. While this particular theory may seem absurd to some in light of modern wisdom, the theory of the born criminal has never been abandoned. For example, there are still those engaged in similar research regarding the structure and composition of the human brain, rather than external human physical characteristics, and its relationship to predicting who is a criminal and who is not. Similarly, geneticists continue to study human genes that may correlate with criminal potential and disposition.

Whitechapel: the forensic pathologists

In the late 1800s, statistical comparison and classification of shared physical characteristics were not the only way that criminal characteristics were being reasoned and inferred for the purposes of criminal investigation. During the Whitechapel murders in Great Britain in 1888, Dr George B Phillips, the divisional police surgeon, engaged in a more direct method of inferring criminal characteristics: not by comparing the characteristics of statistically averaged offenders, but by carefully examining the wounds of a particular offender's victims. That is to say, by examining the behavior of a particular criminal with a particular victim. For example, Dr Phillips noted that injuries to one of the Whitechapel victims, Annie Chapman, indicated what he felt was evidence of professional skill and knowledge in their execution. In particular, he was referring to the postmortem removal of some of Annie Chapman's organs, and what he felt was the cleanliness and preciseness of the incisions involved.

Whatever the basis of his inferences regarding the unknown offender's level of skill, and despite its dubious accuracy given modern understanding of wound patterns in the case, the implication of this type of interpretation is very straightforward. As Dr Wynne E Baxter, Coroner for the South Eastern District of Middlesex, stated to Dr Phillips during a coroner's inquest into the death of Annie Chapman, 'The object of the inquiry is not only to ascertain the cause of death, but the means by which it occurred. Any mutilation which took place afterwards may suggest the character of the man who did it.' Behavior, they understood, suggests personality characteristics.

At the time of the Whitechapel murders, coroners were required to inquire into the nature, character and size of all wounds and to document them thoroughly (though not necessarily by photograph). This practice speaks to the value placed, even then, on what today may be referred to as wound pattern analysis. It is extremely unlikely that the Whitechapel murders are the first crimes where those involved in the investigation utilized wound pattern analysis. However, the investigation does offer some of the earliest written documentation of the types of inference drawn from violent, aberrant, predatory criminal behavior by those involved in criminal investigations. Today, wound pattern analysis is still considered by many to be an important part of the criminal profiling process.

Profiling in the 1900s: the psychiatrists

In the twentieth century, the work of American psychiatrist Dr James A Brussel of Greenwich Village, New York was considered by many to have advanced the thinking behind the criminal profiling process significantly. As a clinician, his approach to profiling was diagnostic. Dr Brussels's method included the diagnosis of an unknown offender's mental disorders from behaviors evident in the crime scenes. He would infer the characteristics of that unknown offender from his own experiences with patients who shared similar disorders. He also subscribed to the opinion

that certain mental illnesses were associated with certain physical builds, not unlike the theories of Lombroso a century before. As a result, an unknown offender's physical characteristics were included in Dr Brussel's profiles of unsolved cases.

During the 1940s and 1950s, a 'mad bomber' terrorized the city of New York, setting off at least 37 bombs in train stations and theaters. Dr Brussel made an analysis of the case. He determined that the man responsible for the crimes was paranoid, hated his father, was obsessively loved by his mother, lived in the state of Connecticut, was of average build, middle-aged, foreign born, Roman Catholic, single, lived with his mother or sister, and wore a buttoned, double-breasted suit. When the police arrested George Metesky for the bombings in 1957, Brussels's profile was thought by many to be very accurate, right down to the suit.

Between June 1962 and January 1964, 13 sexual strangulation-homicides were committed in the city of Boston, Massachusetts and were determined to be related. Traditional investigative efforts by law enforcement to develop viable suspects and identify the 'Boston strangler' were unsuccessful. A profiling committee composed of a psychiatrist, a gynecologist, an anthropologist and other professionals was brought together to create what was referred to as a 'psychiatric profile' of the type of person responsible for the killings.

The profiling committee came to the opinion that the homicides were the work of two separate offenders. They based this opinion on the fact that one group of victims was of older women, and that one group of victims was of younger women. The profiling committee also felt that the psychosexual behavior differed between the victim groups. They felt that the older victims were being strangled and murdered by a man who was raised by a domineering and seductive mother, that he was unable to express hatred towards his mother, and as a result directed it towards older women. They felt that he lived alone, and that if he was able to conquer his domineering mother he could express love like normal people. They were further of the opinion that a homosexual male, probably an acquaintance, had killed the younger group of victims.

Not everyone agreed with the profiling committee. Law enforcement invited Dr Brussel into the investigation in April 1964, in the hope that he would provide them with the same types of insights that helped solve the mad bomber case in New York. Dr Brussel disagreed with the profiling committee. He was of the opinion that the homicides were the work of a single offender. But by then the killings had stopped and the profiling committee was disbanded.

In November 1964, Albert DeSalvo was arrested for the 'green man' sex crimes. He confessed to his psychiatrist that he was the Boston strangler. Since he 'fitted' the profile that Dr Brussel had provided law enforcement so closely, they identified him as their offender and closed the case without filing charges. In 1973, while serving a sentence for the green man crimes, a fellow inmate stabbed DeSalvo to death in his cell. As DeSalvo was never tried for, or convicted of, the crimes committed by the Boston strangler, neither profile has ever been validated.

The FBI: a multidisciplinary approach

It was also during the 1960s that the American Howard Teten began to develop his approach to criminal profiling while still at the San Leandro Police Department in California. His inspiration for the work included Dr Hans Gross, an Austrian magistrate. Teten studied under, and was further inspired by, Dr Paul Kirk, the internationally renowned criminalist, Dr Breyfocal, the then San Francisco Medical Examiner, and Dr Douglas Kelly, a psychiatrist noted for his work in the Nuremberg war trials. They had been his instructors at the School of Criminology, at the University of California, Berkeley during the late 1950s. A multidisciplinary understanding of forensic science, medicolegal death investigation and psychiatric knowledge became the cornerstone of Teten's investigative skills early on, and shaped his approach to criminal profiling.

As a special agent for the Federal Bureau of Investigation, Howard Teten initiated his profiling program in 1970. He taught profiling as an investigative aid to be used in conjunction with other investigative tools. Teten taught his first profiling course, called 'Applied Criminology', to the FBI National Academy in 1970. Later that same year, Teten rendered his first actual profile as an FBI Agent in Amarillo, Texas.

In 1970, Teten also teamed with Pat Mullany, then assigned to the New York Division of the FBI, to teach the abnormal psychological aspects of criminal profiling. Mullany and Teten taught together at several other schools around the country during the next year while Mullany was stationed in New York. They would dissect a crime, Mullany would talk about a range of abnormal behavior, and Teten would discuss how that behavior could be determined from the evidence found at the scene.

In 1972, the US government opened the new FBI Academy, and Teten requested that Mullany be transferred there. Shortly after coming to the new FBI Academy, Teten and Mullany applied their concepts to the first FBI hostage negotiation guidelines. In 1974 and 1975, Mullany negotiated several major hostage situations successfully with these new tech-

niques. These adaptations, based upon criminal profiling techniques, were the first to be taught to all FBI negotiators. They were later modified and expanded, facilitating the creation of what would become known as the FBI's Behavioral Science Unit (BSU).

The FBI's BSU was restructured in the late 1990s, and currently operates as part of the National Center for the Analysis of Violent Crime (NCAVC) at the FBI Academy in Quantico, Virginia.

The FBI, however, is not the only agency in the United States, or in the world for that matter, with a dedicated profiling unit. In the United States, profiling units of various capabilities can be found in a number of federal agencies, as well as numerous state and large city law enforcement agencies. Internationally, law enforcement agencies with dedicated profiling units can be found in countries such as Australia, Canada, England, the Netherlands and South Africa. Additionally, there are a number of independent individuals and organizations involved in criminal profiling around the world. Independent profilers tend to be professionals with investigative, forensic and/or behavioral training, education and experience who work with law enforcement agencies and attorneys.

Profiling Methodology

Contrary to a popular belief that is informed by both fictional and nonfictional media sources, there is nothing mystical or magical about the criminal profiling process. It does not involve psychic ability, and does not rely upon the supernatural in any way. Still thought of as more of an art than a science, criminal profiling is founded in the psychological, sociological and forensic sciences.

Purposes of criminal profiling

A number of separate, related professions have been involved in criminal profiling either directly or indirectly. The education and experience of criminal profilers over the years has reflected this diversity. This may be due to the fact that profiling has not traditionally been a career in itself for most of those who profile, but rather a single tool among many that are available for analyzing, understanding and/or investigating criminal behavior.

Regardless of the background of the person rendering the criminal profile, there are generally two separate applications of the profiling process, each with its own set of goals and guidelines. These are investigative applications and courtroom applications. The profiling process itself, however, is generally the same, regardless of the intended application.

The investigative applications of criminal profiling involve the analysis of behavior of unknown offenders for known crimes. This takes place before a suspect has been generated, and often after all other investigative leads have been exhausted.

Investigative goals

- To reduce the viable suspect pool in a criminal investigation, and to develop the most viable suspects from that reduced pool.
- To assist in the linkage of potentially related crimes by identifying unique crime scene indicators and behavior patterns (i.e. modus operandi and signature aspects).
- To assist in assessing the potential for escalation of nuisance criminal behavior to more serious or more violent crimes (i.e. harassment, stalking, voyeurism).
- To provide investigators with relevant leads and strategies
- To help keep the overall investigation on track and undistracted.

The courtroom applications of criminal profiling involve the analysis of behavior of known crimes for which there is a suspect or a defendant. Profiling takes place in preparation for trial (criminal, penalty and/or appeal phases of the trial are all appropriate times to use profiling techniques).

Courtroom goals

- To assist in the process of evaluating the nature and value of forensic evidence relating to victim and offender behavior.
- To help gain insight into offender fantasy and motivations.
- To help gain insight into offender state of mind through behavior before, during or after the commission of a crime (i.e. levels of planning, remorse, precautionary acts, etc.).
- To assist in the linkage of crimes by virtue of unique signature behaviors.

Profiling methods

There have been essentially two ways of reasoning in the criminal profiling process. One has been called inductive, and refers to the comparative process that is most like the development of psychological syndromes and average offender types. The other has been called deductive, and refers to a forensic evidence-based investigative process of reasoning about the behavior patterns of a particular offender.

Inductive profiling Inductively rendered profiles are often equated with psychological syndromes. It is important to understand that a syndrome is essentially a

cluster of related symptoms. The clinical diagnosis of a syndrome involves comparing an individual's behavior with the behaviors of others in similar circumstances that have been studied in the past. The mental health community christened the first psychological syndrome in 1980, when posttraumatic stress disorder (PTSD) was officially recognized by inclusion in the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III). Since that time, the investigative term 'profile' has slowly become synonymous with the psychological term 'syndrome' in the eyes of the legal community. This has occurred because of the way inductive criminal profiles have been constructed and presented in court.

An inductive criminal profile is a set of offender characteristics that are reasoned, by statistical inference, to be shared by offenders who commit the same type of crime. An inductive profile is best understood as an average. It is the product of statistical or comparative analysis and results in educated generalizations, hence the use of the term inductive. It is very much like a syndrome in its construct.

Inductively rendered criminal profiles include those based upon both formal and informal studies of known, incarcerated criminal populations. They also include criminal profiles based solely upon practical experience, where anecdotal data are recalled by an investigator and used to form the basis for reasoning regarding specific offender characteristics. The result of this process is a boilerplate, an average profile consisting of generalizations about the characteristics of a certain type of criminal.

Example of inductive reasoning A 24-year-old white female is raped in her apartment on the first floor. A given study of 20 serial rapists indicates that they tend to attack within their own age range and do not normally attack outside of their own race. Therefore it is inferred that the offender responsible for the unsolved rape is likely a white male and approximately 24 years old.

A major advantage of an inductive profile is that one can be brought to bear in an investigation relatively quickly, and does not require the application of specialized knowledge. A major disadvantage is that is tends to be brief and nonspecific.

Deductive profiling A deductive criminal profile is a set of offender characteristics that are reasoned from the convergence of physical and behavioral evidence patterns within a crime or a series of related crimes. It is referred to as a deductive process because it involves reasoning in which conclusions about offender characteristics follow from the arguments presented in terms of physical and behavioral evidence.

According to those who use deductive techniques, the information used to infer a deductive criminal profile includes forensic evidence, victimology and crime scene characteristics.

Forensic evidence A competent forensic analysis must be performed before this type of profiling can begin, to insure the integrity of the behavior and the crime scene characteristics that are to be analyzed. All of the behavior between the victim and the offender that is to be used in the profile must be established through crime reconstruction efforts. This can be done using accepted reconstructive techniques, such as wound pattern analysis, bloodstain pattern analysis, bullet trajectory analysis, entomological analysis, or the results of any other physical evidence-based forensic analysis performed. It can also include the use of corroborated victim and witness statements.

Victimology This is the thorough study and analysis of victim characteristics. The characteristics of an individual offender's victims can lend themselves to inferences about offender motive, modus operandi and the determination of offender signature behaviors. Part of victimology is risk assessment. Not only is the profiler interested in the amount of risk a victim's lifestyle places them in routinely, but the amount they were in at the time of the attack, and the amount of risk the offender was willing to take to acquire them.

Crime scene characteristics These include, among many others, method of approach, method of attack, methods of control, location type, primary scene/secondary scene determination, nature and sequence of sexual acts, materials used, verbal activity, elements of planning and precautionary acts. Crime scene characteristics are determined from the forensic evidence and the victimology. In cases involving a related series of offenses, such as in serial rape or serial homicide, crime scene characteristics are determined individually and analyzed as they evolve, or fail to evolve, over time. An offender's crime scene characteristics help the profiler discriminate between modus operandi behaviors and signature behaviors, as well as lending themselves to inferences about offender state of mind, planning, fantasy and motivation.

Example of deductive reasoning A 24-year-old white female is raped in her apartment on the first floor at 12.30 a.m. on a weeknight. She lives alone. She had all of the doors locked; the rapist gained access through an unlocked window. He wore gloves, wore a mask, bound her with duct-tape he had brought with him, used a knife to control her

throughout the attack, and during the rape wore a condom that he had brought with him. He also called her by name several times during the attack, but she did not recognize his voice. After raping the victim, he made a quick search of the apartment and robbed her of credit cards, jewelry and some cash she had hidden in a drawer. Therefore:

- By virtue of the gloves and the mask it can be inferred that the rapist was concerned about leaving behind fingerprints and revealing his identity to the victim.
- By virtue of the condom it can be inferred that the offender planned to commit the rape.
- By virtue of the skill evidenced in the robbery by the swift search for valuables and the precautionary acts, it can be inferred that the offender may have committed this type of crime in the past.
- By virtue of the condom it can be inferred that the offender may have committed rape or sexual assault in the past.
- By virtue of the fact that the offender used her name and wore a mask it can be inferred that the offender is someone that she might recognize, an acquaintance perhaps, or that the offender has preselected the victim as a target and has been surveilling her for some time in advance.

A major advantage of the deductive profiling process is that it provides very pointed insight relevant to a specific pattern of criminal behavior. Another major advantage of the deductive profiling process is that it can take into account offender behavior as it changes over time, because the profile can be refined as new behavior becomes evident throughout an investigation. Major disadvantages of the deductive profiling process include the fact that it takes a great deal of time and effort, and that it requires a great deal of multidisciplinary training and knowledge on the part of the profiler. It is, furthermore, only as reliable as the information that it is based upon.

See also: **Serology**: Bloodstain Pattern Analysis. **Causes of Death**: Scene of Death, **Crime-scene Investigation and Examination**: Scene Analysis and Reconstruction, **Psychology and Psychiatry**: Overview.

Further Reading

Block EB (1958) *The Wizard of Berkeley*. New York: Coward-McCann.

Burgess A and Hazelwood R (1995) *Practical Aspects of Rape Investigation: A Multidisciplinary Approach*, 2nd edn. New York: CRC Press.

Burgess A, Douglas J and Ressler R (1988) *Sexual Homicide: Patterns and Motives*. New York: Lexington Books.

Burgess A, Burgess A, Douglas J and Ressler R (1997) *Crime Classification Manual*. San Francisco: Jossey-Bass.

Canter D and Alison L (1997) *Criminal Detection and the Psychology of Crime*. London: Ashgate.

Geberth V (1996) *Practical Homicide Investigation*, 3rd edn. New York: CRC Press.

Gross H (1924) *Criminal Investigation*. (English translation.) London: Sweet and Maxwell.

Groth AN (1979) *Men Who Rape: The Psychology of the Offender*. New York: Plenum Press.

Holmes R and Holmes S (1996) *Profiling Violent Crimes: An Investigative Tool*, 2nd edn. Thousand Oaks, CA: Sage.

Thornton J and Kirk P (eds) (1974) *Crime Investigation*. 2nd edn. New York: Wiley.

Turvey B (1999) *Criminal Profiling: An Introduction to Behavioral Evidence Analysis*. London: Academic Press.

Vollmer A (1949) *The Criminal*. New York: Foundation.

CRIMINALISTICS

D M Gialamas, Los Angeles County Sheriff's Department, Los Angeles, CA, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0464

Criminalistics in the Forensic Sciences

In its broadest sense, forensic science is defined as the application of the sciences as it pertains to legal matters or problems. These matters may be involved with criminal laws, such as a case involving a perpetrator accused of murder; with civil laws, such as a case involving the determination of liability in an industrial accident; or with governmental regulations, such as the random drug testing of government employees. Several specialties, or disciplines, can be found within the forensic sciences. These include areas such as medicine/pathology, criminalistics,

engineering, odontology, entomology, anthropology and many others. Though the terms 'criminalistics' and 'forensic science' may have different meanings in different parts of the world, the information presented will consider criminalistics as a discipline within the forensic sciences.

In the United States, criminalistics is the broadest subdivision found in the forensic sciences. Criminalistics, which is best defined by the California Association of Criminalists (CAC), is the profession and scientific discipline directed toward the recognition, identification, individualization and evaluation of physical evidence by application of the natural sciences to law–science matters. A criminalist uses the scientific principle of chemistry, biology and physics to elicit information from crime scenes and physical evidence. Through the application of the scientific method using these natural sciences, the evaluation of evidence can be accomplished in a clear, unbiased and accurate manner. Adherence to the scientific method directs the forensic scientist to advocate the truth on behalf of the evidence, not for a particular side.

History of Criminalistics

The history of criminalistics does not have a single person to credit for its inception. In fact, much of the technology in criminalistics is borrowed from other sciences and applied to legal matters. There were many contributors to the birth of the field through the mid-1800s to the early 1900s. Sir Arthur Conan Doyle's Sherlock Holmes is often credited as the fictional father of criminalistics, using methods in criminalistics long before the science was recognized and accepted. Mathieu Orfila is credited as the father of forensic toxicology. He is also credited with being the first expert witness in a criminal trial in the 1840s. In the 1880s Alphonse Bertillon created the first classification system attempting to achieve personal individualization, called anthropometry, and Hans Gross coined the term *Kriminalistik*. Shortly thereafter, in the 1890s, Francis Galton published his study on fingerprints. Around the 1910s several contributors were added to history: Landsteiner discovered blood groups (i.e. A, B, H); Leone Lattes obtained blood types from dried blood; Calvin Goddard published his work on firearms comparisons; and Albert Osborn undertook document examinations. Then, in 1920, Edmond Locard postulated his 'exchange principle', a fundamental contribution to the field upon which many of the specialties in criminalistics are based. Locard's police laboratory in Lyon, France, was so successful that it gave the needed impetus to the formation of police laboratories and crime laboratories in Europe and the United States. In fact, the first crime laboratories were opening in the late 1920s in the United States. In the 1930s police science and criminalistics were emerging in academia, and in the late 1940s the school of criminology was formed at the University of California at Berkeley, headed by Paul Kirk. As new technology was emerging in the 1960s, England's Home Office created the Central Research Establishment, the first forensic research center in the world. The 1970s, 1980s and 1990s saw an explosion of information and analytical methods, as the world's technological advances and instrumentation were (and still are) improving.

Physical Evidence

Physical evidence can be anything, from massive objects as large as a building to submicroscopic molecules of a fleeting vapor. Some say that physical evidence is a silent witness. It can help show that a crime has been committed (*corpus delecti*) and give insight to the perpetrator's method of operation (*modus operandi*). Physical evidence can provide investigative leads, provide a connection between suspect(s) and victim(s), help identify and individualize persons or objects involved, and assist in substantiating or disproving witness statements.

The premise for our work in criminalistics is based on Locard's exchange principle. It states that whenever two objects come into contact with one another, there is always a transfer of material across the contact boundaries. In the 1950s, Paul Kirk added that no criminal can perpetrate a crime without leaving evidence behind or taking evidence away. Science and technology, therefore, are the limiting factors in the detection of transfers of evidence. The role of the criminalist is to recognize and collect these evidence exchanges at the scene of the crime and, through the rigorous examination of physical evidence in the laboratory, help make the facts of the case clear for an investigator, judge or jury.

An Introduction to the Disciplines in Criminalistics

Criminalistics is unique among the forensic sciences in that it is the only specialty to have a number of seemingly unrelated disciplines within it. This article will take an introductory look at these disciplines, briefly describing the types of evidence encountered and the methods used in each area. For further details regarding each of the disciplines and their examination processes, the reader is referred to the appropriate areas in the encyclopedia.

Crime scene processing

Truly, the most challenging and the most important aspects of any physical evidence examination begin at the scene of the crime. The recognition, documentation and collection of physical evidence are crucial steps needed to elicit information from physical evidence. These tasks can be accomplished with the criminalist or crime scene investigator who has special knowledge, skills and abilities. The fundamental tasks involved in any crime scene include securing the crime scene, properly searching and documenting the scene, and recognizing, collecting and packaging physical evidence. Because each crime scene is unique, criminalists need good communication with the investigators and flexibility with the crime scene environment in order to accomplish these tasks in an expedient, yet efficient and competent manner. No amount of work or dedication in the laboratory can ever substitute for a poorly processed crime scene.

Some crime scenes require further knowledge, skills and abilities in particular subject areas in order to provide crime scene reconstructions. These scenes may include clandestine laboratory investigations where illicit and dangerous drugs are manufactured, crime scenes with extensive bloodstain patterns that must be interpreted, arson and explosive crime scenes with their massive and intense destruction, and scenes involving firearm trajectories with several firearms being discharged. Together with the evidence examination results, the criminalist can connect the information and render an opinion regarding the events of the crime.

Forensic photography

Photography is an art known to many, but few recognize that photography in the forensic sciences requires not only the fundamental photographic skills but also the understanding and appreciation of physical evidence and the need to properly document it. Forensic photographers take advantage of the properties of light, special filters and film emulsions available to them to create a permanent record of physical evidence. Many large laboratories have a staff of photographers with the equipment and studio necessary to properly document evidence. For some evidence categories, photography is the only means in which evidence can be collected and preserved.

The forensic photographer may also process the film and provide enlargements for court displays, maintain archive negative files for cases, and provide photographic equipment repair and technical advice. He or she is called out to crime scenes to document evidence when the photographic expertise of the crime scene investigator is limited. Forensic photographers will examine and provide expert testimony in cases where the evidence item is exposed film.

The new technology of imaging is creating exciting changes for the forensic photographer, who may also be working with computerized formats for image capturing in addition to emulsion film. Forensic imaging may soon be the title that replaces forensic photography. For example, the Los Angeles County Sheriff's Department is converting from obtaining booking photographs on conventional film emulsion to taking them on a computerized, digital format.

Fingerprints

Fingerprint evidence is a very well known evidence category that can individualize evidence to a single person. Fingerprint examiners provide individualizations and eliminations of latent friction ridge prints obtained from objects at the crime scene to known persons. Although the evidence items are routinely fingerprints, they may also include palm, toe or foot prints. During the analysis, the fingerprint examiner compares the class (i.e. the print pattern, such as arch, loop or whorl) and minutiae (i.e. the presence of ridge features within the pattern, such as bifurcations, islands, ridge endings and their spatial relationships) of the objects of interest. Individualization is achieved when a sufficient number of minutiae are present between the comparative prints. Currently, there is some disagreement among the experts as to what constitutes a sufficient number of comparative points. Some argue that a minimum number must always be found, whereas others offer that the number will depend on the minutiae present.

Special methods are available for the development of latent fingerprints, including fingerprint powders, chemical processing and laser or alternative imaging techniques. Larger law enforcement agencies may also have automated fingerprint identification systems (AFIS) that automate fingerprint searches against computer databases. A qualified fingerprint examiner compares the possible candidates selected in the search to determine if an individualization can be made.

Drug analysis

A drug can be defined as a natural or synthetic substance that will affect a physiological or psychological change in the human body. These substances, prior to the introduction into the human body, can be found in many forms, including plant materials, powders, tablets, capsules, liquids and gases. Many drugs are controlled or regulated by laws, which can

depend on their medicinal characteristics and their potential for abuse. The drug analysis case load in many laboratories is generally high. Well over 50% of all cases submitted to crime laboratories in the United States involve drug, alcohol or other toxicological analysis.

The task of the criminalist is to isolate and identify the unknown substance and determine whether it is a dangerous or illicit drug. The preliminary examination of materials may include visual examination, magnified examination through a stereobinocular microscope, and chemical spot tests, which indicate the presence of a drug type depending on a color change. Confirmation of preliminary results is usually obtained from microcrystalline tests using a polarized light microscope, Fourier transform infrared (FTIR) absorption spectroscopy or gas chromatography with mass spectrometry (GC–MS).

Toxicology and blood alcohol

Toxicology, meaning the 'study of toxins' in root form, is closely related to drug analysis. Drug analysis concerns natural or synthetic substances before introduction into the body, whereas toxicology concerns these substances and their metabolites after introduction into the body. Toxicology can, in part, provide investigative information regarding driving-while-impaired cases, contribute to random drug testing of employees and assist pathologists with cause of death determination. Blood alcohol determination, although a part of toxicology, has become such an important analysis in the eyes of society that many laboratories in the United States have separated this analytical area from other toxicological methods.

The task of the toxicologist or blood alcohol analyst is to isolate and identify alcohol (specifically ethanol), illicit and prescription drugs, poisons and other toxins in blood, urine and tissue samples. Alcohol can also be obtained from human breath samples. The analysis determines the qualitative (presence) and quantitative (amount) properties of drugs and toxins in the body. The preliminary examinations may include extraction and purification methods followed by radioimmunoassay (RIA), thin-layer chromatography (TLC) and ultraviolet absorption (UV) spectroscopy. Modern screening techniques incorporate enzyme immunoassay methods with little or no extraction or purification of the sample. Currently, the application of GC–MS is the standard used for confirmation and quantitation. An evaluation of the results may include expert opinions regarding mental and/or physical impairment of the tested subject while using the detected substances or their metabolites.

Forensic biology

Historically, there has been a distinction between conventional forensic serological methods and deoxyribonucleic acid (DNA) methods. Conventional serological methods have been limited in the forensic sciences because of the difficulty, or even inability, of many testing procedures to successfully yield results on small, dried stains. Advances in DNA methods can now reliably test smaller and more degraded samples and have replaced many of the conventional serological methods. Today these examinations fall under the generalized heading of forensic biology, and the original task of identification and individualization of body fluids and tissues remains the focus.

The most common samples tested in the laboratory are blood, semen and vaginal fluids. Other body fluids, such as feces, milk, saliva, sweat, tears and urine, and tissues, such as brain, bone, skin and muscle, may also challenge the criminalist in the identification and individualization of persons of interest. The identification process begins with the visual observation of the physical properties, such as color, solubility and taction; chemical color or luminescent tests; ultraviolet (UV) or alternate light source (ALS) fluorescence or dampening; and stereobinocular, brightfield, phase contrast and other microscopical examinations with or without staining techniques. Once confirmation of a body fluid or tissue is made, it is followed by species identification to determine whether the sample is of human or animal origin.

Individualization of a bodily fluid or tissue is performed by a comparison of genetic markers from the evidentiary item with the exemplar reference sample(s). Conventional serological methods include agglutination techniques (e.g. ABO typing) and electrophoretic methods comparing polymorphic enzymes and proteins. Modern DNA methods utilize advances in gene mapping to compare regions of interest of DNA strands. Restriction fragment length polymorphism (RFLP), polymerase chain reaction (PCR), amplified fragment length polymorphism (AmpFLP), mitochondrial DNA (mtDNA) and other methods are used to individualize DNA fragments.

Trace evidence

Trace evidence is a unique discipline in criminalistics, as it has within itself many specialties. Trace evidence is the only discipline in the forensic sciences that includes a wide range of examined materials that are unrelated. The methods and techniques used in trace evidence examinations can be broken down into three evidence categories: patterned evidence, comparative evidence, and physical and chemical properties evidence.

Patterned evidence is the most powerful of the trace evidence categories. It includes the identification and individualization of footwear, tire track, glove, fabric and clothing impressions and physical matches of cut, broken or torn items. During the analysis, the criminalist compares the class and accidental characteristics of the objects of interest. Individualization is achieved when a sufficient number of accidental characteristics are present between the comparative items. The examination of patterned evidence includes visual and stereomicroscopic examinations. First, the criminalist compares the class characteristics, which include the size, shape and design of the manufacturer's pattern, and, if appropriate, the general wear pattern, on the evidence item with the exemplar. Then the criminalist locates and compares the accidental characteristics, which include the size, shape and spatial relationship of the points of interest. These points may be nicks and gouges on a shoe outsole or tire tread, a jagged edge from a torn item, or any other unique feature. Methods of enhancement may be needed to resolve sufficiently the detail of a patterned impression. For items to be individualized, the evidence items and the exemplars must be examined and compared in their true, life-sized forms; nonlife-sized photographs or other reproductions without a scale are unsuitable for comparison.

Comparative evidence is the next category and includes the identification and the attempt toward individualization of items. The types of evidence encountered by the criminalist include natural and synthetic textile fibers, fabrics and ropes; human and animal hairs; soil samples; and a variety of glass and paint sources. Comparative evidence analysis begins with visual and stereormicroscopic preliminary examinations. Chemical spot tests may also be used in the preliminary examination of paint. Further examination is performed using various types of comparison microscopy (e.g. brightfield, polarized, differential, phase contrast and fluorescent), FTIR, ultraviolet-to-visible range (UV-Vis) spectrophotometry, GC, GC–MS, pyrolysis–GC or GC–MS, TLC, X-ray fluorescence (XRF) and scanning electron microscopy with energy dispersive X-ray (SEM-EDX) microanalysis. Often this type of evidence corroborates or refutes investigative information. It is only under rare circumstances that these types of evidence categories yield an individualization.

Physical and chemical properties evidence involves only the identification or detection of a substance in evidence items. These trace evidence areas include gunshot primer residue (GSR) analysis, mace and pepper spray detection, exploding money-dye pack detection, fire debris analysis, explosives residue analysis and explosive device reconstruction, automotive lamp on–off determination, and unknown material identification. The methods of examination are similar to the chemical, microscopic and instrumental methods used in the analysis of comparative evidence.

Toolmarks and firearms

Toolmarks and firearms analysis involves the identification and individualization of indentations and striations on surfaces. Firearms evidence comparisons are a specialized area of toolmark analysis. Toolmark analysis may involve the comparison of screwdrivers and crowbars to window pry marks, and pliers and bolt cutters to cut items such as padlocks and cables. The firearms examination unit routinely compares striations on bullets and bullet jacketing and striations and indentations on cartridge cases to specific firearms. In the absence of a suspect firearm, the examination of these items may provide information regarding the possible types of firearm capable of producing the marks on the items. The comparison begins with the reproduction of the evidence mark with the suspect tool. This is easier for firearms, as the marks are replicated by firing a cartridge in the firearm and recovering the casing and bullet. Toolmarks, on the other hand, can pose a real challenge to the criminalist who is attempting to replicate the exact angle and direction of the tool relative to the surface, as a change in either can alter the characteristics of the indentation or striation. The laboratory comparison of toolmarks and firearms evidence is done using visual, stereomicroscopic and brightfield microscopy methods.

A firearms criminalist also examines firearms and their components (e.g. silencers) for proper functioning, an important aspect in accidental discharge cases. Muzzle-to-target distance determinations, traditionally involving clothing items, can be estimated by the firearms examiner with the proper firearm and ammunition. In cases of serial number restoration, the criminalist takes advantage of chemistry and metallurgy in an attempt to recover obliterated or otherwise removed serial numbers.

Questioned documents

A questioned document may be any object upon which there are character markings that are in doubt. This usually involves written and machine-printed materials on paper but may include any markings made on an object, such as paint or blood written on a wall or floor. For written documents, the examiner is concerned with the identification and individualization of handwriting and handprinting. In order to ascertain the authenticity or source of the

questioned document, examiners may need to identify and decipher indented writings; identify and elicit information from altered documents, including additions, deletions, erasures, burning and fabrications; and identify and compare inks from multiple types of writing instrument and printed documents.

Machine-generated documents have changed rapidly with advancing technology, offering more challenges to the document examiner. Today's technology has provided document examiners with greater resources for examination and comparison. It has also added questioned documents, such as credit card evidence, which includes card embossing, magnetic strip decoding and duplicate copy signatures; computer generated documents, which includes printed material and data files; and photocopied evidence.

As with trace evidence, preliminary examinations for the questioned document begin with visual and stereomicroscopic inspection. Advanced techniques in microscopy and infrared (IR) and alternate light source (ALS) imaging are heavily relied upon, especially for altered documents. Image capturing through digital cameras and importation into graphics programs for enhancement are also utilized. Chemical methods, such as TLC, and instrumental methods, such as spectroscopy, are used in the comparison of inks and dyes.

Reporting and court testimony

Once the work of a criminalist is completed, whether at the crime scene or in the laboratory, the results of the examinations must be communicated. This is accomplished through written reports. The laboratory reports should include the date, requesting agency and case/file number, the items received, the examination methods, the results of the analysis, and the opinion and signature of the examiner. The issued written report becomes a legal document that can be introduced and used as evidence in the courtroom.

On occasion, testimony will accompany the written report in the courtroom. Testimony by an expert witness can provide the judge and/or jury with an in-depth explanation of the examiner's findings and opinions. Of course, the breadth of subject matter may be limited by the judge's or attorney's questions. The criminalist must qualify as an expert witness each time he or she testifies through a process of *voir dire*.

Concluding Remarks

The future of criminalistics will be an exciting one to watch. With the aid of the spectacular advances in science, criminalistics will continue to improve its use of technology and provide more answers from physical evidence examinations to previously unresolved questions. As the technology changes, so, too will the quality of the profession. Laboratory accreditation, which began as a voluntary program, is now an essential criterion for an operational crime laboratory in the United States. Certification of criminalists, whether general or in a specialty area, is currently voluntary and will likely shift to one that becomes essential in the years ahead.

Criminalistics is without a doubt the most diverse specialty in the forensic sciences. The variety of physical evidence items examined in criminalistics makes it a challenging and rewarding professional occupation for many.

See also: **Administration of Forensic Science**: An International Perspective. **Alcohol**: Blood. **Basic Principles of Forensic Science**. **Serology**: Bloodstain Pattern Analysis. **Crime-scene Investigation and Examination**: Recording; Collection and Chain of Evidence; Recovery of Human Remains; Packaging; Preservation; Contamination; Fingerprints; Suspicious Deaths; Major Incident Scene Management; Scene Analysis and Reconstruction; Criminal Analysis. **Document Analysis**: Handwriting; Analytical Methods; Forgery/Counterfeits; Ink Analysis; Document Dating. **Evidence**: Classification. **Expert Witness**: Qualifications and Testimony. **Fingerprints (Dactyloscopy)**: Visualization; Sequential Treatment and Enhancement; Identification and Classification; Standards of Proof; Automated Methods, including Criminal Record Administration; Chemistry of Print Residue. **Firearms**: Weapons, Ammunitions and Penetration; Range; Residues. **History**: Forensic Sciences; Crime-scene Sciences; Fingerprint Sciences. **Pattern Evidence**: Footwear Marks; Shotgun Ammunition on a Target; Tools; 'Plastic' Bag Striations; Vehicle Tire Marks and Tire Track Measurement; Serial Number. **Pharmacology**. **Photography and Digital Imaging**: Overview.

Further Reading

DeForest PR, Gaensslen RE and Lee HC (1983) *Forensic Science: An Introduction to Criminalistics*. New York: MacGraw-Hill.

Eckert WG (ed) (1997) *Introduction of Forensic Sciences*. Boca Raton, FL: CRC Press.

Fisher BAJ (1993) *Techniques of Crime Scene Investigation*, 5th edn. Boca Raton FL: CRC Press.

Hilton O (1993) *Scientific Examination of Questioned Documents*. Boca Raton, FL: CRC Press.

Kirk P (1953) *Crime Investigation*. New York: Interscience.

Saferstein R (ed) (1982) *Forensic Science Handbook*, vol. I. Englewood Cliffs, NJ: Prentice-Hall.

Saferstein R (ed) (1988) *Forensic Science Handbook*, vol. II. Englewood Cliffs, NJ: Prentice-Hall.
Saferstein R (ed) (1993) *Forensic Science Handbook*, vol. III. Englewood Cliffs, NJ: Prentice-Hall.
Saferstein R (1995) *Criminalistics: An Introduction to Forensic Science*, 5th edn. Englewood Cliffs, NJ: Prentice-Hall.
Tuthill H (1994) *Individualization: Principles and Procedures in Criminalistics*. Salem, OR: Lightning Powder Company.

D

Dactyloscopy *see* **Fingerprints (Dactyloscopy)**: Automated Methods, including Criminal Record Administration; Chemistry of Print Residue; Identification and Classification; Sequential Treatment and Enhancement; Standards of Proof; Visualization.

Databases *see* **Deoxyribonucleic Acid**: Databanks.

Daubert *see* **Legal Aspects of Forensic Science**.

Debit Cards *see* **Forgery and Fraud**: Payment Cards.

Defence Wounds *see* **Clinical Forensic Medicine**: Defense Wounds.

Dentistry *see* **Odontology**.

DEOXYRIBONUCLEIC ACID

Contents

Basic Principles
Databanks
Future Analytical Techniques
Mitochondrial Deoxyribonucleic Acid
Parentage Testing
Polymerase Chain Reaction
Polymerase Chain Reaction-Short Tandem Repeats
Restriction Fragment Length Polymorphism
Significance
Statistical Analysis

Basic Principles

M S Schanfield, Analytical Genetic Testing Center, Inc, Denver, CO, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0467

Introduction

Forensic science is applied science. That is to say that the scientific methodologies used in forensic science were developed by biologists, chemists and geneticists and then taken over by forensic scientists to help them

solve problems. Research in the 1950s, 1960s and 1970s identified polymorphic enzymes such as acid phosphatase (ACP, referred to as erythrocyte acid phosphatase (EAP) by forensic scientists), adenosine deaminase (ADA), adenylate kinase (AK), carbonic anhydrase (CA2, only useful with samples of African American origin), esterase D (ESD), glyoxalase (GLO), peptidase A (PEPA, primarily of use with samples of African American origin), and phosphoglucomutase (PGM1) and proteins such as group specific component (GC, now known as Vitamin D binding globulin (VDBG)), hemoglobin (HB, primarily used with samples of African American origin), haptoglobin (HP), the immunoglobulin allotypes (GM and KM), and transferrin (TF). These markers were used in the late 1970s and early 1980s by forensic science in the United States and abroad to individualize bloodstains. Although some of these markers did not last long in bloodstains they made it possible to individualize bloodstains with greater than a 99% certainty. Semen stains, saliva and urine had relatively few markers that could be detected, so it was often difficult to provide information in cases with these types of evidence. The ABO blood groups (Secretors), Lewis blood group, immunoglobulin allotypes, and the enzyme PGM1 were the markers routinely used in these cases, with PEPA and GLO occasionally useful.

In 1980 with the identification of the first hypervariable DNA polymorphism (D14S1), detected by restriction length polymorphism technology (RFLP), the door was opened to the possibility that DNA technology could be applied to forensic evidence. In the next several years, the search for new markers led to the identification of many forensically useful markers, detected by RFLP, some of which are still used routinely in forensic DNA testing. In the mean time hypervariable minisatellite regions were found which identified many genetic regions at one time (multilocus probes). The term 'DNA fingerprinting' was used to describe these bar code like patterns. Although these regions proved to be highly informative for parentage testing, they did not have the sensitivity needed for forensic testing. Though multilocus probes are used outside of the United States for paternity testing and some forensic applications, they were never widely accepted in the United States.

At the same time as the revolution in RFLP was beginning in the mid 1980s, the development of modern polymerase chain reaction (PCR) tests occurred. The role of the polymerase enzymes in copying DNA had been known since the 1970s. The use of high temperature Taq polymerase allowed for the automation of thermal cycling and the introduction of modern PCR. Tests were developed to identify human leukocyte antigens (HLA) for the transplantation community. The first marker to be developed was to the HLA region called DQα (now called DQA1). This became the first PCR-based test to be used forensically.

In the mid 1980s the only DNA RFLP testing on forensic evidence available in the United States was through two commercial laboratories. Initially DNA RFLP testing was only done on cases that could not be resolved by non-DNA enzymes and protein testing. The initial success rate was low, approximately 10%, largely due to the nature of the evidence tested. As the success rate for testing sexual assault evidence improved so did the demand for testing. The FBI began to develop procedures of RFLP typing and establishing a standardized system for publicly funded crime laboratories in the United States. Similar work was going on in Canada, the UK and Europe. Initially, different restriction enzymes were used. The Lifecodes Corporation used the enzyme Pst I; Cellmark used the enzyme Hinf I; and the FBI chose the enzyme Hae III. The United States and Canada standardized using a forensic RFLP-system based on the enzyme Hae III whereas England and Europe standardized by using the enzyme Hinf I. In the late 1990s all forensic DNA testing using RFLP is done using either the restriction enzyme Hae III or Hinf I.

What is DNA?

DNA stands for deoxyribonucleic acid. It is the biological blueprint of life. DNA is made up of a double-stranded structure consisting of a sugar (deoxyribose) and phosphate backbone, cross-linked with two types of nucleic acids referred to as purines (adenine and guanine) and pyrimidines (thymine and cytosine) (**Fig. 1**). The cross-linking nucleic acids always pair a purine with a pyrimidine, such that adenine always pairs with thymine and guanine always pairs with cytosine.

DNA can be found in several areas of a cell. The majority of DNA is located in the nucleus (**Fig. 2**) organized in the form of chromosomes (22 pairs of autosomes and a set of sex chromosomes (X and Y)). Each nucleated cell normally has 46 chromosomes that represent the contribution from both parents. In the formation of gametes (eggs and sperm) one chromosome of each pair is randomly separated and placed in the gamete. The separation of chromosomes is referred to as segregation. The transmission of half of our chromosomes to our children in the form of gametes is the basis of Mendelian inheritance. This DNA is referred to as nuclear or genomic DNA. With the exception of identical twins, no two people share the same genomic DNA sequence.

Another source of DNA is found in the mitochondria in the cytoplasm of cells (Fig. 2). Unlike nuclear

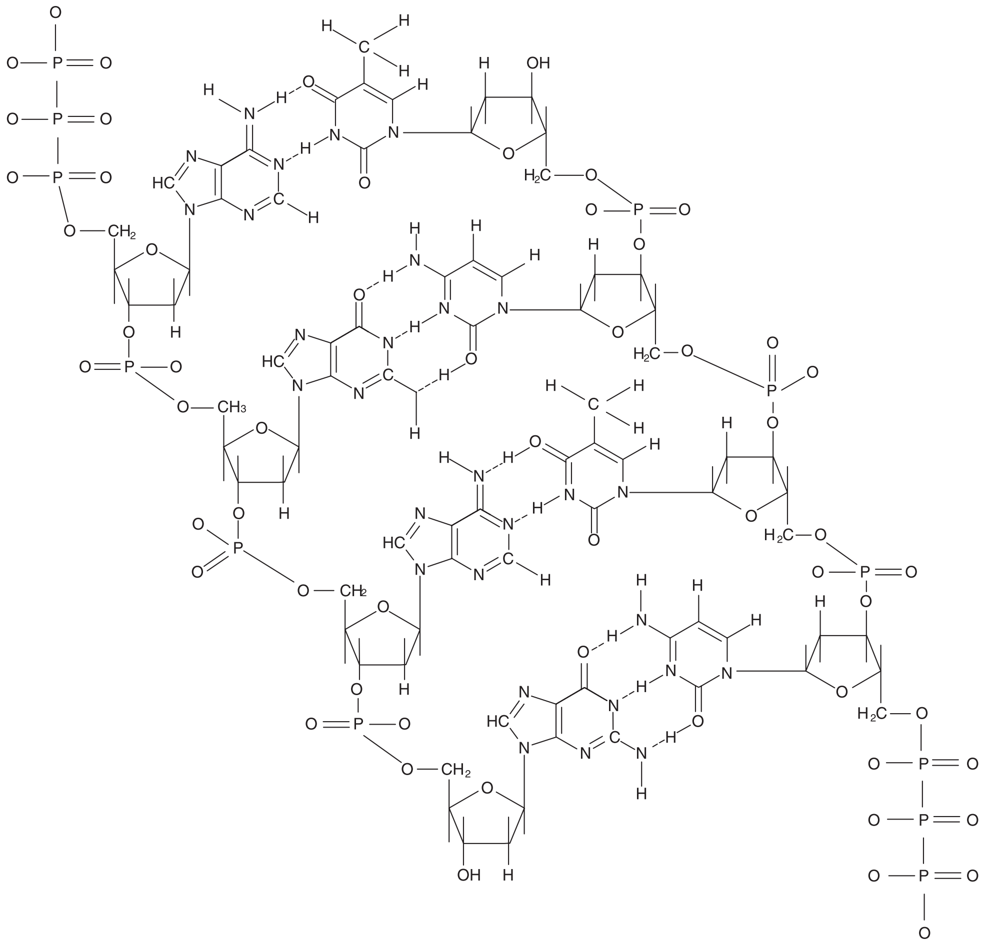


Figure 1 Molecular structure of DNA. From top to botttom: Adenine-Thymine, Guanine-Cytosine, Adenine-Thymine and Guanine-Cytosine.

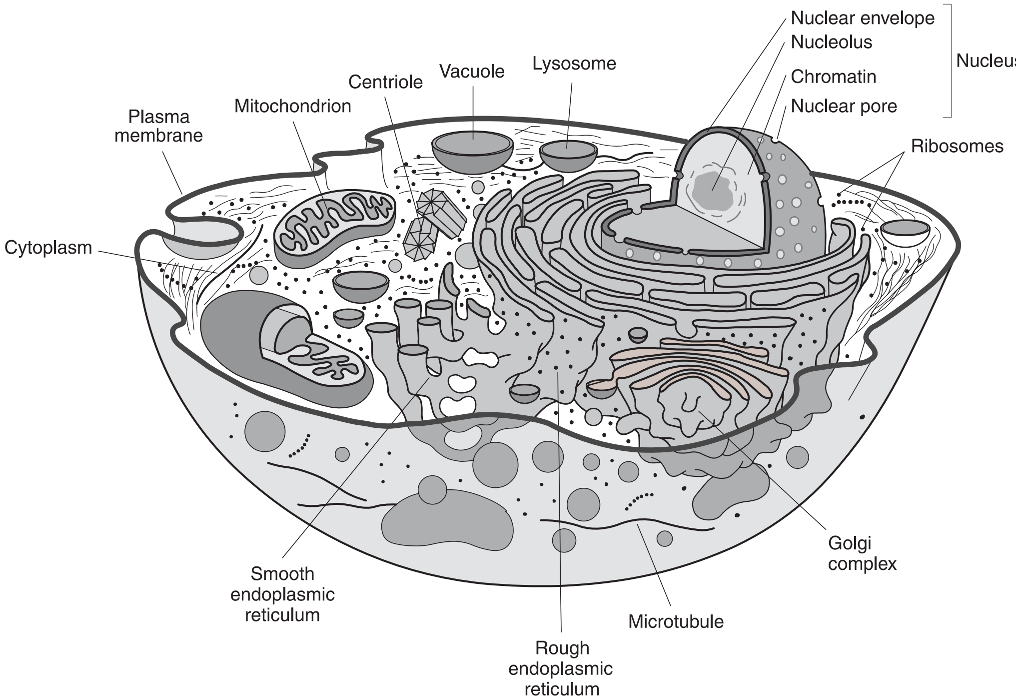


Figure 2 A generalized eukaryotic cell showing the organization and distribution of organelles as they would appear in transmission electron microscope. The type, number and distribution of organelles is related to cell function.

DNA, which only has two copies of each genetic region, mitochondrial DNA is involved in energy production within the cell and can have between 100 and 10 000 copies per cell. Structurally, instead of a linear arrangement of DNA within chromosomes, mitochondrial DNA has a circular structure. Mitochondrial DNA is inherited from the mother because it is found in the cytoplasm which comes from the egg (ova).

Where is DNA Found?

Nuclear or genomic DNA is found in all nucleated cells as well as in the reproductive cells (eggs and sperm). The amount of DNA we can expect to find in different cells and types of evidence is found in **Table 1**. DNA has been successfully obtained from blood and bloodstains, vaginal and anal swabs, oral swabs, well-worn clothing, bone, teeth, most organs and to some extent urine. It is less likely that DNA will be obtained from some types of evidence than others. Blood or semen stains on soil and leather are historically not good sources of evidenciary DNA. Saliva *per se* has few nucleated cells, but, beer and wine bottles, drinking glasses, beer cans, soda cans, cigarettes, stamps and envelope flaps have all been found to provide varying amounts of DNA.

How much DNA is Needed for Forensic Testing?

The amount of DNA needed to perform testing depends on the technology used. RFLP technology usually needs at least 50 ng of intact high-molecular-weight DNA. In contrast PCR-based testing can use as little as 500 pg. Most PCR-based tests are set up to use between 1 and 10 ng of genomic DNA.

Destruction of DNA

Biological materials are affected by their environment. Enzymes lose activity over time and type of storage conditions. DNA has been found to be relatively robust when it is in the form of dry stains. Initial environmental studies indicated some of the limitations of DNA based on the material it is deposited upon and the environmental conditions. Environmental insult to DNA does not change the results of testing, you will either obtain results, or if the DNA has been too badly affected by the environment (i.e. the DNA is degraded) you do not get RFLP results. One report on the success rate of obtaining RFLP results noted that depending on the substrate or condition of the stain, results were obtained between 0 (carpet stains or putrefied samples) and 61.5% (scrapped dried stains) of the time with an average of 52% for the 100 items of evidence tested. Thus, the material that the DNA is deposited on and the degree of further insult can markedly affect the ability to obtain RFLP DNA results.

All the published studies on environmental insult were done on prepared dried stains. Since biological fluids are liquid the effects of ultraviolet radiation on liquid DNA have been evaluated. The results of exposing 100 μl samples of a standard DNA solution to fluorescent light in the laboratory, a UV germicidal light (254 nm), mid day sunlight in January, and early sunset light in January in 15 min increments, up to 1 h are presented in **Fig. 3**. There is a linear decrease in high-molecular-weight DNA with the UV germicidal light, such that after an hour about 96% of the high-molecular-weight DNA has been lost. Even in the weak mid day light in January, over 60% of the high-molecular-weight DNA was lost. In contrast, the fluorescent lighting in the laboratory and the after sun set light had no effect on the amount of high-molecular-weight DNA. This was not a rigorous experiment, but the effects are dramatic enough to demonstrate the effect of ultraviolet light exposure to DNA before stains dry.

Table 1 DNA content of various tissues

1 sperm	3 pg
1 cell	6 pg
1 shed hair	1 ng[a]
1 plucked hair	300 ng[b]
1 drop of blood	1500 ng

[a] The success rate for PCR on shed hairs is 30–50%, so this average is optimistic.
[b] There is a great deal of variation among hair roots. Fine blond hair tends to have much less DNA, whereas a dark hair with large roots has more.

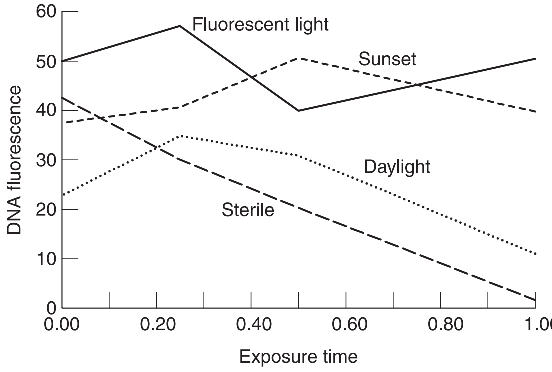


Figure 3 Plot of DNA concentration (flourescence) over time, after exposure to different light sources.

Extraction of DNA

As stated previously DNA exists inside cells. Because most evidence is in the form of dry stains, the DNA must be removed from the stain before it can be tested. The process of removing DNA from the cells on the evidence and dissolving it is referred to as extraction. There are several procedures available for removing DNA from evidence so that it can be used. They are referred to as either 'organic' extraction or 'nonorganic' extraction based on the nature of the chemicals used. Further, there are two special types of extraction. The first, called differential extraction, was developed for sexual assault evidence to separate the cells that come from the victim (epithelial cells from the vagina, rectum or mouth) from those of the perpetrator (male sperm cells). The second method is a specialized 'nonorganic' extraction using Chelex beads. Chelex beads can only be used when PCR-based DNA testing is going to be used. The basic DNA extraction procedures, whether organic or non-organic, can be adapted for special circumstances such as hair or tissue.

Chloroform–phenol extraction

This is the oldest procedure available for extracting DNA from blood and it has been extended to include hair, tissue and semen stains. The basic procedure consists of opening up cells with a buffer and an enzyme, usually protease K, and then denaturing and separating the proteins from the DNA. The latter part is done using a mixture of chloroform (chloroform:isoamyl alcohol 24:1) and phenol (buffered). The phenol–chloroform mixture denatures proteins liberated by the first stage. The major disadvantage of this procedure is the fact that phenol–chloroform is a hazardous waste and could theoretically pose a risk to pregnant employees. A modern protocol for phenol–chloroform extraction of various types of evidence can be found in the literature.

Nonorganic extraction

In nonorganic extraction the hazardous phenol–chloroform protein denaturation step is replaced by a salting out of proteins. This allows for the same chemistry to be used for the initial phase of DNA extraction, and replacement of the hazardous elements of the procedure with a nonhazardous alternative. The salting out procedure has several advantages over the phenol–chloroform extraction. The first is that instead of having two liquid phases (organic and nonorganic) that can occasionally trap the DNA in the wrong (organic) phase, by precipitating the proteins (e.g. the proteins become insoluble and become a solid), there are liquid and solid phases with the DNA only in the liquid phase (nonorganic). The second advantage is that the hazardous phenol–chloroform is replaced with a harmless salt solution. Comparison of the organic and nonorganic procedures for blood and semen indicate that the nonorganic extraction is on the average as good as or better than organic extraction, whether quantitated by yield gel or slot blot (**Table 2**).

Either method of DNA extraction described above can be used for both RFLP or PCR-based DNA testing. Organic DNA extraction is widely used in laboratories doing criminal case work whereas nonorganic DNA extraction is widely used in laboratories performing paternity testing, research and diagnostics. On a worldwide basis nonorganic DNA

Table 2 Comparison of organic and nonorganic extraction of DNA from blood and semen stains

Quantitation method	*Blood*		*Semen*	
	Organic	*Nonorganic*	*Organic*	*Nonorganic*
Yield gel				
Mean	185 ng	258 ng	175 ng	207 ng
N	21	8	22	8
P	0.054		0.122	
Slot blot				
Mean	515 ng	908 ng	627 ng	1175 ng
N	22	8	27	8
P	0.022		0.008	

Data taken from Tables 1 and 2 of Laber *et al* (1992). Evaluation of four deoxyribonucleic acid (DNA) extraction protocols for DNA yield and variation in restriction fragment length polymorphism (RFLP) sizes under varying gel conditions. *Journal of Forensic Sciences* **37**: 404–424. Differences in means were tested by Kruskall-Wallace nonparametric analysis of variance, H statistic with 1 d.f., uncorrected *P* values presented.

extraction is the more prevalent. With the shift to PCR-based testing this choice in extraction is becoming increasingly common.

Chelex extraction

In 1991 a method of DNA extraction was described that was specifically aimed at the extraction of small amounts of dilute DNA for PCR-based testing using Chelex beads. The method is simple, relatively fast and biohazard free. It is widely used by forensic laboratories doing PCR-based typing which has increased the number of laboratories using nonorganic, biohazard free DNA extraction. The only limitations of Chelex extraction is that it produces a dilute solution of DNA that may need to be concentrated before it can be used with some of the newer high resolution PCR-based typing systems.

Quantitation of DNA

Yield gel quantitation

Whether RFLP- or PCR-based testing is performed it is necessary to know how much DNA is present. One of the earliest methods of quantitating small amounts of DNA is the use of a yield gel. A small gel is made using a salt solution to carry electrical current and a supporting medium made of agarose (a complex carbohydrate made from seaweed). Much like gelatin, the agarose is dissolved in water that is heated to near boiling, the liquid is cooled slightly and poured into a forming tray. A plastic comb or former with rectangular teeth is placed in the liquid agarose. Once the agarose gels, the comb is removed leaving behind rectangular wells in the agarose gel. The DNA to be tested is mixed with loading buffer, and placed in the wells. Loading buffer is a mixture of a large amount of sugar and dye. The high concentration of sugars makes the mixture heavier than the salt solution so that the DNA sinks to the bottom of the well. The dye allows the migration of the DNA to be monitored. As the agarose was melted in water containing salt, when an electrical current is applied to the gel, electricity flows through the gel because of the salt and moves (migrates) from the negative electrode (cathode), toward the positive electrode (anode). Since all DNA has a negative charge, and was placed in the wells at the cathodal end of the gel, the negatively charged DNA will migrate out of the wells toward the positive end of the gel. If the DNA is broken into pieces that are different sizes, the smaller pieces will move through the gel faster than the larger pieces and will be separated based on size. This process of separating DNA using an electric current is called electrophoresis, which simply means separation (phoresis) by means of electricity (electro).

Since DNA is colorless it is not possible to see the DNA after it has been separated without the use of special dyes that bind to it. One of the earliest dyes used was ethidium bromide which fluoresces pink when bound to double-stranded DNA and exposed to ultraviolet light. **Fig. 4** is an ethidium bromide-stained yield gel. To quantitate the amount of DNA in the DNA extracts, a set of DNA quantitation standards are placed on the gel. By visual comparison of the unknown DNA with the known DNA the amount of DNA can be approximated. This test provides information about the relative amount of DNA and whether it is degraded (i.e. the DNA is broken down so that different size pieces of DNA are present). It does not indicate if the DNA is human, however, since all DNA will fluoresce. Thus the DNA present may be bacterial as well as human DNA. For RFLP testing the total amount of DNA in the sample is the important determinant of how the samples migrate in the gel. Therefore, yield gel electrophoretic quantitation of DNA is an appropriate method. Yield gel quantitation of DNA for RFLP testing was considered to be such an integral part of quality assurance that it was included in the National Institute of Standards, Standard Reference Material 2390, 'DNA Profiling Standard.'

As with the extraction of DNA using the organic method, ethidium bromide is potentially hazardous

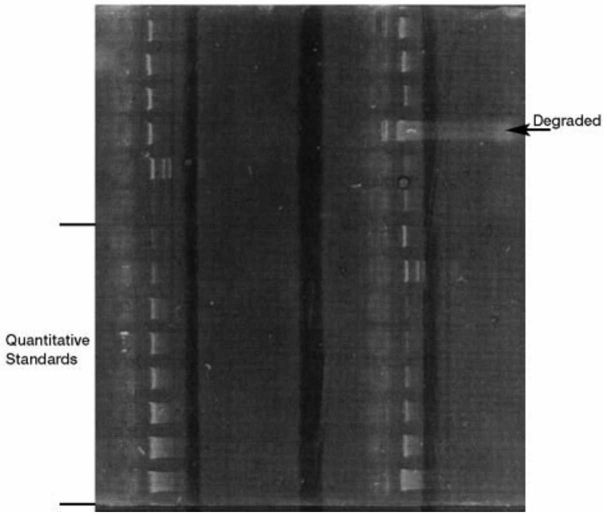


Figure 4 Ethidium bromide stained yield gel. Bottom left samples are quantitative standards. Other samples represent various samples of DNA. Upper right sample is degraded DNA.

because the dye is associated with an increased cancer risk. Although ethidium bromide is still widely used for the identification of DNA it is currently being replaced by a new dye called Sybr® green which is much less carcinogenic and can detect smaller amounts of DNA than ethidium bromide.

Slot blot quantitation

In contrast to RFLP, for PCR-based testing, the amount of human DNA and not the total amount of DNA is an important determinant in how likely it will be to obtain results. A slot blot does not rely on electrophoresis to separate the DNA but rather on the ability of denatured (separated DNA strands) DNA to bind to homologous complementary sequences. The ability to quantitate human DNA requires sequences of DNA that are common in the human genome so that a single DNA sequence can recognize them and bind them. The repeated DNA sequence called D17Z1 is the basis for all human DNA slot blot quantitation systems. There are several of these procedures commercially available. In one of the most widely used tests, the quantitation requires that denatured DNA is applied to a membrane using a slotted plastic apparatus. The denatured DNA binds to the membrane. The membrane is exposed to a solution of denatured DNA fragments that recognizes a repeating sequence of human or primate DNA. Pieces of DNA that recognize a specific region of DNA are referred to as a 'probe'. The probe will bind to complementary DNA fragments stuck on the membrane. The probe has an indicator attached to it so that the binding of the DNA to the probe can be detected. The unbound probe is washed off and the probe is detected using either chemicals that change color (colorimetric detection) or chemicals that give off light (chemiluminescent detection). To be able to quantitate the amount of human DNA present, standards with different amounts of human DNA are also placed on the membrane. A series of known DNA quantitation standards are included in the blot so that it is possible to determine the approximate amount of DNA bound to the membrane by visual comparison to the known standards. More precise quantitation can be obtained by scanning the membrane with a scanning densitometer and determining the amount of color associated with each band. Most forensic laboratories use visual comparison.

See also: **Deoxyribonucleic Acid**: Restriction Fragment Length Polymorphism; Polymerase Chain Reaction; Polymerase Chain Reaction-Short Tandem Repeats; Future Analytical Techniques; Significance; Databanks; Mitochondrial Deoxyribonucleic Acid.

Further Reading

Adams DE, Presley LA, Baumstark AL *et al.* (1991) Deoxyribonucleic acid (DNA) analysis by restriction fragment length polymorphism of blood and other body fluid stains subjected to contamination and environmental insults. *Journal of Forensic Science* **36**:1284–1298.

von Beroldingen CH, Blake ET, Higuchi R, Sensabaugh G and Erlich H (1989) Application of PCR to the analysis of biological evidence. In: Ehrlich HA (ed.) *PCR Technology: Principles and Applications for DNA Amplification*, pp. 209–223. New York: Stockton Press.

Haugland RP (1996) *Handbook of Fluorescent Probes and Research Chemicals*, 6th edn, pp. 144–156. Eugene, OR: Molecular Probes.

Jeffreys A, Wilson V and Thein SL (1985) Hypervariable 'minisatellite' regions in human DNA. *Nature* **314**: 67–73.

Laber TL, O'Connor JM, Iverson JT, Liberty JA and Bergman DL (1992) Evaluation of four deoxyribonucleic acid (DNA) extraction protocols for DNA yield and variation in restriction fragment length polymorphism (RFLP) sizes under varying gel conditions. *Journal of Forensic Sciences* **37**:404–424.

Maniatis T, Fritsch EF and Sambrook J (1982) *Molecular Cloning: A Laboratory Manual.* New York: Cold Spring Harbor Laboratory.

McNally L, Shaler RC, Baird M, Balazs I, Kobilinsky L and De Forest P (1989) The effects of environment and substrata on deoxyribonucleic acid (DNA): the use of casework samples from New York city. *Journal of Forensic Science* **34**:1070–1077.

Miller SA, Dykes DD and Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research* **16**:1215.

Mullis KB, Faloona FA, Scharf SJ, Saiki RK, Horn GT and Erlich HA (1986) Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction. *Cold Spring Harbor Symposium on Quantitative Biology* **51**:263–273.

Shows TB, McAlpine PJ, Bouchix C *et al.* (1987) Guidelines for human gene nomenclature. An international system for human gene nomenclature (ISGN, 1987). *Cytogenetics and Cell Genetics* **46**:11–28.

Walsh PS, Metzger DA and Higuchi R (1991) Chelex® 100 as a medium for simple extraction of DNA for PCR based typing from forensic material. *BioTechniques* **10**:506–513.

Walsh PS, Varlaro J and Reynolds R (1992) A rapid chemiluminescent method for quantitation of human DNA. *Nucleic Acids Research* **20**:5061–5065.

Waye JS and Willard HF (1986) Structure, organization and sequence of alpha satellite DNA from human chromosome 17: evidence for evolution by unequal crossing-over and an ancestral pentamer repeat shared with the human X chromosome. *Molecular and Cellular Biology* **6**:3156–3165.

Wyman A and White R (1980) A highly polymorphic locus in human DNA. *Proceedings of the National Academy of Sciences (USA)* **77**:6754–6758.

Databanks

B D Gaudette, Royal Canadian Mounted Police, Ottawa, Canada

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0474

Introduction

DNA analysis was undoubtedly the most important development in forensic science in the twentieth century. Its ability to analyze small and environmentally challenged samples and to accurately establish their origins with a high degree of certainty was exploited to advantage by forensic scientists in the latter years of the century. However, forensic DNA analysis has another property that may eventually be even more important. Just as with fingerprints, the results of forensic DNA analysis can be stored in a databank. This can lead to many unique and important developments. The findings from most disciplines of forensic science come into play late in the investigative process where a suspect has been determined through other means and the police are looking for evidence that would either confirm or deny the person's association with a crime and that could assist in any prosecution. DNA databanks are important because they assist at the beginning of the investigative process, enabling police to determine or eliminate suspects in the absence of other evidence. In addition, DNA databanks play another unique role. Even before the crime occurs, they can make a contribution which is of infinite value in economic and social terms, namely crime prevention. This can arise in many ways. First, it has been shown that violent offenders often commit several crimes (six or more according to one survey cited by the FBI) before they are finally apprehended. By making it possible to identify the perpetrator after their first or second offense, DNA databanks spare many potential victims. Secondly, by increasing the certainty with which dangerous violent offenders can be apprehended and successfully prosecuted, DNA databanks help to keep such people 'off the streets'. Finally, the knowledge that they are very likely to be apprehended if they re-offend acts as a deterrent to some released convicts whose DNA information is in a databank.

The benefits of DNA databanks must be weighed against potential costs. Benefits are seldom achieved without causing some harm, and DNA databanks are no exception. In addition to the financial costs of their establishment and maintenance, DNA databanks also engender significant privacy concerns. Accordingly, protection of privacy should be one of the pillars in the design of DNA databanks and the legislative framework surrounding them. An equally important design and legislative consideration is to maximize their efficiency and effectiveness as an investigative tool. Databank design and privacy concerns will be discussed, as will be the interrelationship of fingerprint and DNA databanks, benefits to police and society, and reaction of offenders. We will begin with a look at the history of DNA databanks.

History of DNA Databanks

On 1 August 1986, Dawn Ashworth was murdered in Leicestershire UK. The police involved in the search for the person who murdered her (and another victim from 1983) put a newly discovered tool, forensic DNA analysis, to work in their investigations. They created an informal DNA database by collecting blood samples from all the males in the town. As events unfolded, the database included all the males in town but one; Colin Pitchfork, the person eventually charged with the murder, had someone else pose for him in giving a blood sample. When that person later spoke about this in a pub, Colin Pitchfork was found out and the world's first DNA database led, albeit indirectly, to his arrest.

From the time they first started conducting DNA testing in the mid 1980s, the London Metropolitan Police Laboratory started banking DNA information on an *ad hoc* basis. By the early 1990s this databank had become a key weapon in the detection of sexual crimes. However, in 1992 this databank was ordered to be destroyed when a legal action was brought before the European Commission on Human Rights under the privacy provisions of the European Convention.

The US Armed Forces Institute of Pathology established the first formal DNA databank in 1992. This databank, which was created to identify service men missing in action, proved successful in identifying casualties from the 1991 Desert Storm operation. By sharing their experience and through sponsorship of annual conferences on DNA databank and repositories beginning in 1992, the Armed Forces Institute of Pathology made a major contribution to development of forensic DNA databanks.

In the late 1980s, several states in the USA became the first jurisdictions in the world to pass DNA databank legislation. Other US states soon followed suit and the FBI developed the Combined DNA Index System (CODIS) to enable state databanks to communicate and share data with each other. Before long the first cold hit in a DNA databank occurred in Minnesota in 1991. A rape murder was solved by a

match between DNA from blood and semen found at the crime scene and a databank sample from a person imprisoned for a burglary. The police went on record as saying that the case may never have been solved without the DNA databank. The wisdom of developing CODIS was shown in 1994 when the first cross-jurisdiction case was solved. By early 1999, all 50 American states had DNA databank legislation and a federal DNA databank was under development. Two factors initially limited the success of American DNA databanks. There was a lack of resources to analyze all samples collected from offenders and a long-established policy of not analyzing samples from cases without good suspects existed in many forensic laboratories. The DNA Identification Act of 1994 and other legislation provided funding for DNA databanks and forensic laboratories to overcome these obstacles as long as states comply with specified quality assurance standards, submit to external proficiency testing, and limit access to DNA information.

In 1994, legislation enabling establishment of a DNA databank was passed in the UK and in 1995 the Forensic Science Service opened the world's largest and most successful DNA databank. The tremendous success of this databank spurred many other countries to pass DNA databank legislation. Numerous other countries around the world now have operational DNA databanks.

Design of DNA Databanks

In order to withstand court challenges and be viable in the long term, DNA databanks require enabling legislation. Such legislation will reflect the political, social, economic and constitutional environment of the jurisdiction in which it is enacted. Since no two jurisdictions are exactly alike for all factors, differences will naturally occur in the resulting legislation. This, in turn, will lead to differences in the design of the databanks. The following discussion highlights the various options available.

Indices

Two indices form the core of virtually all DNA databanks. The first of these, usually termed the crime scene index, contains DNA information obtained from bodily substances connected with crime scenes. The second, often known as the offenders index, contains DNA information from known samples of bodily substances taken from criminal offenders. DNA databanks in some jurisdictions contain additional indices such as missing persons DNA, or anonymous samples used to generate population statistics and establish evidential value.

Who is required to provide samples?

Depending on the jurisdiction, only a small subset of offenders may be included in the convicted offenders index (e.g. only those persons convicted of major sexual offences), there may be a long list of offenses for which persons convicted are required to give samples, or (as is the case in the UK), all persons charged with any criminal offence could potentially have their DNA information entered into a databank. In most jurisdictions, the process of conviction (or charge) is sufficient of itself to require an offender to provide a sample. Some jurisdictions, on the other hand, require that a special application be made to the presiding judge. Canada has two classes of offenses; *primary offenses* (most major crimes of violence) for which those convicted are compelled to provide samples, and *secondary offenses* (less serious offenses) where judicial discretion applies. Legislation in many jurisdictions includes provision for retroactive collection from persons convicted of offenses prior to enactment of the legislation and/or for retrospective collection from those persons who committed crimes before the legislation came into force but were convicted subsequently.

What samples are taken

Blood, being the most reliable bodily substance for forensic DNA analysis, is the most common sample collected. In many jurisdictions, liquid blood is used since tubes of blood are often collected routinely as part of the prison intake process. Other jurisdictions collect bloodstains with a finger prick, which is less intrusive and eliminates the requirement to have medical professionals involved in the collection process. This can save a considerable amount of money and is more convenient with respect to collection from convicted persons who receive a sentence that does not involve prison. Finger prick samples can now be collected on FTA®-treated paper. This greatly simplifies subsequent DNA extraction, and improves health and safety by immobilizing blood-borne pathogens. In some jurisdictions collection of blood is felt to be too intrusive. Two other sample types are then available – plucked hairs and buccal swabs. Some jurisdictions permit collection of two or three sample types and encourage collection of a second sample type in case the first does not provide satisfactory results due to improper collection or storage.

What is retained in the databank?

The large databank created by the US Armed Forces Institute of Pathology is unique in that only bodily substances (liquid blood samples) are stored. The bank does not store DNA information since analysis

is conducted only when unidentified remains are found in relation to a military incident. All other forensic DNA databanks store results of DNA analysis of collected samples of bodily substances. In nearly all jurisdictions, the bodily substances themselves are also stored. This enables the databank to remain current in the event that a major change in analytical technology renders existing DNA information obsolete. However, the bodily substances themselves engender far greater privacy concerns than the limited DNA profiles used for forensic identification purposes. Accordingly, a few jurisdictions do not permit retention of samples after analysis; DNA databanks in such jurisdictions consist of DNA information only.

Analytical methodology

The first DNA databanks were based on restriction fragment length polymorphism (RFLP) analysis. Newer databanks use polymerase chain reaction (PCR) analysis of short tandem repeats (STRs). Many older databanks are in the process of conversion to the newer methodology as it is faster and more amenable to automation. Several databanks are increasing the use of robotics to automate parts of the analytical process. It is important that sufficient DNA loci be analyzed to ensure that the databank returns only one match and not several. A North American standard of 13 STR loci has been developed and is presently in use in many countries around the world. Britain and some other European countries use a somewhat different set of loci but there is sufficient overlap between the two standards to enable some sharing of data. In the future, increased standardization is anticipated in response to the increasingly international nature of modern crime.

Statistical evaluation

It is very important to assess the value of a match between the DNA profiles associated with a crime scene and those of a suspect developed through a databank search. This topic has been quite controversial, and there has been a gradual evolution of views as the meaning of a match and the appropriate questions to be answered became better understood. Initially, many people felt that since it is much more likely that a crime scene profile will match with someone in a large databank than it would with any one suspect developed through conventional means, a databank search would drastically reduce the evidential strength against an individual found to match. This was the view taken by the US National Research Council. Their 1992 report recommended that when a suspect is identified through a databank search, additional loci should be analyzed (in both the crime scene and suspect profiles) and that 'only the statistical frequency associated with the additional loci should be presented at trial'. Their 1996 report suggests that evidence obtained from searching a data bank of size n would be n times weaker than the same evidence obtained by other means. There is a logical inconsistency in this view; a match to only one person from a crime scene profile searched against a databank of every person on earth would be extremely probative, whereas the above argument would suggest that it would have no probative value. It has been shown that this problem arises because the approach advocated by the National Research Council addresses the question: 'how likely is it that someone in the databank would match if all were innocent?' rather than the relevant question: 'given that a databank search has been conducted, how strong is the evidence against the unique individual found to match?'. When this more appropriate question is asked, the answer shows that the DNA evidence is actually slightly stronger after a positive databank search than it is when a suspect is developed through other means, and accordingly, there is no need to use additional loci for court purposes. (It should be noted that whereas the DNA evidence is stronger after a databank search, the overall case against the suspect may very well not be; hence an understanding of the problems of incorporating the DNA evidence with the non-DNA evidence in such cases can be particularly important. In addition, police investigating cases in which a suspect is developed only through a DNA databank match need to guard against the biasing effect this information might have on other aspects of their investigation.)

Other design considerations

Some databanks use bar codes to identify samples, thereby providing an additional measure of privacy protection and minimizing the potential for sample mix-up. Some databanks require the convicted offender to provide a fingerprint along with the sample of bodily substance. This then enables fingerprint and criminal history files to be checked to verify the identity of the offender and ensure that an alias has not been used. Many databanks are required to purge samples and DNA information from certain offenders who have been granted a pardon. Use of bar codes or other means of rapidly locating and identifying samples is particularly important in such jurisdictions.

Critical success factors

Regardless of the design of a DNA databank, the following factors are critical to ongoing successful operation. The enabling legislation must be substantial and robust, and it should be continually amended

and updated. For example, the trend in many jurisdictions is to start with a limited number of designated offenses and later, once the databank has established a good reputation and has the ability to expand, to amend the legislation by adding additional offenses. It is extremely important to ensure that all samples authorized by law are indeed obtained and that they are promptly and properly collected. Proper administration and sample tracking must be established with appropriate documentation. Quality control and quality assurance procedures must be put in place and should be reviewed by a recognized accrediting organization. Loci used in ongoing casework must correspond to loci used in the databank. Finally and most importantly, forensic laboratories must have the capacity to examine cases with unknown suspects.

Privacy Considerations

The DNA molecule holds the key to all that makes an individual unique. DNA information can reveal secrets of a person's past (such as who his or her biological parents were), and can predict predisposition to genetic-based diseases. Were such information to fall into the hands of the wrong people, or to be used in the wrong way by otherwise well-meaning people, much harm could ensue. Privacy advocates fear that samples from DNA databanks will be used in research aimed at identifying a 'criminal gene'. They are also concerned that the 'law enforcement purposes' for which forensic DNA databanks were intended will be gradually broadened to include use by immigration authorities, child support enforcement officials and other government agencies, and that this 'surveillance creep' will eventually lead to everyone being required to have their DNA information on file. For these reasons, those concerned about privacy have carefully scrutinized DNA databanks and their enabling legislation. Fortunately, legislators and people responsible for the operation of forensic DNA databanks have generally been cognizant of privacy concerns and have built in numerous safeguards. First and foremost, loci chosen for forensic DNA analysis and databanks are deliberately selected to be from noncoding regions of the DNA molecule with no known direct link to any genetic disease or trait. Secondly, many databanks use bar codes and other means to ensure that there is no direct link between a person's DNA information and their personal identifiers, thereby making it difficult for any one person to illegitimately obtain the information. Third, authorizing legislation often prescribes fines or prison terms to those people whose action or inaction causes improper release of DNA information or to those who tamper with DNA samples. Fourth, some jurisdictions involve privacy advocates in the drafting of legislation and in the oversight of DNA databank operations.

Interrelationship of Fingerprint and DNA Databanks

DNA databanks serve as a complement to, rather than a replacement for, long-established fingerprint databanks. DNA evidence is most commonly associated with crimes of violence although DNA can also often be found in property crimes. Conversely, fingerprints are most commonly found in property crimes but can also often occur in crimes of violence. Instead of having just one identification tool, police investigators now have a pair. If fingerprints are not found at a particular crime scene, the chances are that DNA will be and vice versa. Both fingerprints and DNA databanks can be used to link crimes committed by serial offenders. If a particular crime scene has both DNA and fingerprints, longer linkages can be established, e.g. three crimes linked to a scene through fingerprints can be connected with four others linked through DNA.

Reaction of Offenders

When police obtain a tool as powerful as DNA databanks, criminals can be expected to offer strong reactions. Some of these are positive. Knowledge that their DNA information is in a databank is sufficient to cause some offenders to refrain from criminal activity. There have even been reported incidents of people who were convicted before the date when a DNA databank was established volunteering to give samples as a demonstration of their determination to change their ways. Having their DNA information in the databank also benefits released offenders who can be quickly eliminated as suspects whenever an unsolved crime occurs in their neighborhood, thereby reducing the amount of hassle they receive from police.

However, as might be expected, the reaction of the majority of offenders is negative. Some try to mount court challenges to the authorizing legislation. Since most legislation has been carefully drafted in a consultative manner, such challenges are almost invariably unsuccessful. Offenders must thus adapt to the new reality and find ways of circumventing it. The first attempt at such adaptation occurred when Colin Pitchfork sent someone else to give a blood sample for him. Anecdotal evidence of other adaptations has included an increased use of condoms in sexual

assaults (which brings other benefits to society), and attempts to destroy evidence through burning bodies etc. Some criminals have even learned to use DNA databanks to their advantage by planting someone else's DNA at crime scenes.

Benefits to Police and Society

DNA databanks can help police quickly solve crimes that would previously have taken a long time to solve or may have remained unsolved. They can link crimes within and across jurisdictions, thereby establishing that a serial offender is at work and increasing the chances of eventual solution. They can also save police much time and expense. When investigating unsolved crimes, police no longer need to conduct time-consuming interviews with previous offenders, who can be quickly and effortlessly eliminated by a quick search of the databank. On the other hand, if a match is obtained, this generally provides enough evidence for arrest, thereby eliminating the need to mount expensive surveillance or undercover operations.

Much as DNA databanks benefit police, it is society as a whole which reaps the greatest benefit – reduction in victimization.

See also: **Deoxyribonucleic Acid**: Restriction Fragment Length Polymorphism; Future Analytical Techniques; Parentage Testing; Mitochondrial Deoxyribonucleic Acid; Significance. **Fingerprints (Dactyloscopy)**: Identification and Classification.

Further Reading

Balding DJ (1997) Errors and misunderstandings in the second NRC report. *Jurimetrics* **37**:469–476.

Balding DJ and Donnelly PJ (1996) Evaluating DNA profile evidence when the suspect is identified through a database search. *Journal of Forensic Science* **41**:603–607.

McEwan JE (1995) Forensic DNA data banking by state crime laboratories. *American Journal of Human Genetics* **56**:1487–1492.

McEwan JE and Reilly PR (1994) A review of state legislation on DNA forensic data banking. *American Journal of Human Genetics* **54**:941–958.

Redmayne M (1998) The DNA database: civil liberty and evidentiary issues. *Criminal Law Review* 437–454.

Schneider PM, Rittner C and Martin PD (1997) Special issue: Proceedings of the European Symposium: Ethical and Legal Issues of DNA Typing in Forensic Medicine. *Forensic Science International* **88**:1–110.

Toolis K (1992) Police delete DNA files on rape suspects. *Sunday Telegraph* 3 May.

Wambaugh J (1989) *The Blooding*. New York: Morrow.

Wangstad W (1991) DNA identifies suspect in slaying. *Saint Paul Pioneer Press* 20 December.

Future Analytical Techniques

V W Weedn, Carnegie Mellon University, Pittsburgh, PA, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0471

Introduction

The dawn of forensic DNA typing is generally dated to 1985 when the first article in the scientific literature suggested that highly discriminatory DNA markers could be used to resolve questions concerning evidence source identification. Since then, forensic DNA typing methods and technologies have been evolving rapidly. The advances have been so rapid that laboratory directors have had difficulty in keeping pace and predicting future needs.

The Human Genome Project has ensured revolutionary technological progress in biotechnology, including those that may have forensic application. This continued rapid pace of change has been anticipated by the US Attorney General's Commission on the Future of DNA Evidence and has chosen to dedicate a working group to future DNA technologies. The panel, chaired by Jim Crow PhD, is expected to be published during the year 2000.

This article begins by reviewing the evolution of forensic DNA technology to date and assessing the trends that can be observed.

Evolution of RFLP Testing

Restriction fragment length polymorphism (RFLP) technology was itself first described by Southern in 1975. Sir Alec Jeffreys deserves credit for describing its use with hypervariable loci and recognized its potential for forensic identification purposes. Jeffreys' multilocus probes, which created a bar code-like pattern, gave rise to the term 'DNA fingerprint'. Jeffreys first applied the RFLP testing for judicial purposes in 1985 in the case of a Ghanaian boy immigrating to Great Britain. The first criminal application of DNA testing was also by Jeffreys, in two United Kingdom (Leicester) rape-homicides. These DNA testing efforts began in that case in 1986 but were not concluded for years later.

RFLP testing methodologies have since undergone numerous advances within forensic laboratories. Multilocus probes quickly gave way to single-locus probes, which were known in the scientific literature as VNTRs. It was difficult to interpret statistically the multibanded pattern produced from multilocus probes and from samples containing more than one

source, as in most rape-derived evidence samples. Furthermore, multilocus probes required more sample DNA and were more technically demanding than single-locus probes in the production of consistent results. This was the result of fading of hybridization signal from low homology DNA bands. Lifecodes Inc. and Cellmark Diagnostics were the first laboratories to offer single-locus probe RFLP testing, in 1986 and 1987. The first criminal conviction based on DNA testing was in the 1986 US case of *Florida* v. *Andrews* using single-locus probes. Single-locus probes generally replaced multilocus probes within the forensic service community in criminal casework by the early 1990s.

Other than Professor Jeffreys' laboratory, commercial laboratories (Lifecodes, Forensic Science Associates, Cellmark) were the only laboratories to perform forensic DNA identity testing for the first 3 years of DNA investigations of crime scene evidence. (The first government laboratories to begin casework were the Federal Bureau of Investigation (FBI) in late 1988 and the Virginia Forensic Science Department in 1989.) These commercial laboratories were responsible for early procedures and the choices of probes and enzymes and the validation of their systems. They developed different sets of reagents and touted their differences to attract customers. The need to transfer the technology to government laboratories, and the discomfort of the courts in admitting proprietary testing methods, caused the US forensic community to look towards standardization. Budowle of the FBI, in conjunction with state and local crime laboratories, created the Technical Working Group on DNA Analysis Methods (TWGDAM). This group explored existing DNA technologies, simplified standard test protocols, and standardized the use of *Hae*III (a four-base cutter) as the preferred restriction enzyme within the US. Jeffreys' methods employed the restriction enzyme *Hinf*I (a six-base cutter) and many in Europe continued to use *Hinf*I.

Data quality produced in autoradiograms from the earliest casework improved rapidly within the first few years as the DNA typing community improved their technical proficiency. Incremental improvements continued as manufacturers developed ever-better product lines. Computer imaging systems permitted more uniform readings. Proficiency surveys provided avenues for better interlaboratory comparisons. In recent years, the most important advances were in the development of chemiluminescent and fluorescent labeling systems that obviated autoradiography. Perhaps as important as the technical advances during this time was the development, fine-tuning and general acceptance of statistical interpretation methods.

This early period of evolution can be seen as a time of introduction and dissemination of DNA typing capability into most forensic laboratories. Most crime laboratories had no experience or expertise in molecular biological techniques before the 1990s. In fact, DNA testing by RFLP electrophoretic methods is a natural evolution of the protein electrophoretic biological testing performed for decades before by forensic serologists. Judicial acceptance was accomplished during this time frame. Frameworks for quality assurance guidelines were also implemented during this time.

Several trends in technology had their origins in this period. The community was very sensitive to the courtroom presentation of the data, requiring the production of increasingly clean results. There was an emphasis on discriminatory power, resulting in an increasing number of polymorphic probes. A reduction in the turnaround time of tests also became important. In particular, chemiluminescence began to replace autoradiography to speed turnaround time as well as to avoid the use of radioisotopes.

Evolution of PCR Testing

The origins of the polymerase chain reaction (PCR) technique trace to 1985, and the first court case using PCR testing was by a private California laboratory (Forensic Science Associates) in 1986. Thus, PCR testing began in parallel to RFLP testing. However, PCR-based testing played a distinctly subservient role to RFLP testing over the first decade of its availability. Although there are many reasons for the dominance of RFLP testing in the early years, chief among these was the poor discriminatory power of early PCR-based tests. Nonetheless, due to rapid advances in PCR technologies, PCR-based tests have generally supplanted RFLP testing in forensic laboratories around the world. The US was generally last to abandon RFLP testing, and PCR-based tests predominated in Europe before most other parts of the world.

The first PCR-based tests involved reverse dot blots, where the test depended on the development of a color reaction from probe hybridization on a filter membrane. Such tests analyze sequence-based rather than length-based polymorphisms. The commercial availability of kits meant that laboratories could adapt them easily. These tests have the inherent PCR advantages of rapid testing and exquisite sensitivity. Furthermore, because the targeted loci are very small, they are particularly useful when the DNA specimens are degraded. The binary 'yes/no' discrete results are a desired advantage over the nondiscrete RFLP tests. However, the early tests had

far less discriminatory power than RFLP systems. Furthermore, the interpretation of mixed stain specimens and weak dots are sometimes problematic. High-density line blots were being developed, which involved a large number of genetic systems and achieved reasonable discrimination comparable to RFLP testing, but have been superseded by advances in other PCR-based technologies.

A method using PCR to take advantage of the polymorphic potential of traditional RFLP loci was developed by Jeffreys: termed minisatellite variant repeat (MVR) mapping, it produced a 'digital' typing result. This technology was never adopted by the forensics community. Similarly, other PCR strategies, such as genetic bit analysis, did not catch the attention of the mainstream forensics community.

Amplified fragment length polymorphism (AmpFLP) analysis is an early PCR-based testing methodology that has been employed by forensic laboratories. These tests involve the analysis of length-based polymorphisms, like RFLP tests. Instead of analysis of restriction fragment length polymorphism size differences, however, AmpFLP tests are used to assess amplified fragment length polymorphisms. Conventional PCR amplification will not successfully amplify the large RFLP fragments; much smaller target fragments are necessary. Thus, new polymorphic loci had to be identified to apply this technique. The small fragment sizes necessitated a stiffer electrophoretic medium than agarose; hence the use of acrylamide gels for AmpFLPs. Probe hybridization is obviated by the ability to directly detect amplified fragments, because amplicons will predominate over the specimen DNA. Discrete alleles may be determined unambiguously. However, longer fragments may become degraded and may not amplify well. Preferential amplification is the tendency for PCR to favor the amplification of shorter fragments over larger fragments and is a potentially interfering artifact to guard against.

The desire for smaller target regions resulted in the development of short tandem repeats (STRs). STRs are a species of AmpFLPs in which the core repeat units are 3–7 bp in length. It should be noted, however, that compound and complex repeat sequences occur. Original AmpFLPs, such as D1S80, contained long tandem repeats in which the core repeat units were larger than STRs. AmpFLPs and STRs were first used in forensic casework for the identification of remains in the Persian Gulf War in 1991.

STR analysis is less susceptible to DNA degradation than other AmpFLPs. Also, preferential amplification is much less problematic with the shorter STR fragments than with other AmpFLPs. So-called 'stuttering' is a PCR artifact of smaller STR systems, particularly trinucleotide systems, but can be minimized by judicious selection of STR primers and amplification conditions. Most forensic STRs contain tetranucleotide repeat units. CSF1PO, THO1, TPOX and vWF (CTT, CTTv) were the most extensively used of the early STR systems. Some interest has recently been expressed in somewhat larger STRs (pentameric repeats) to eliminate stutter and thus resolve issues of mixtures. Large multiplexed sets of STRs permit discriminatory power which rivals that of RFLP. The applicability of STR systems to automated instrumentation platforms may be among their most important attributes.

The Forensic Science Service in the United Kingdom pioneered efforts to create robust and powerful multiplexed sets of STRs (quadruplex, SGEM, etc.) and have since applied them to casework and databasing applications. Many of these sets were adopted by other laboratories.

In the US, the FBI sponsored an STR working group which has advocated the use of 13 core STR loci performed in two amplification reactions for submission into the CODIS database. Manufacturers (PE Biosystems, Promega) have accommodated the community by developing kit assays which respond to this need. These STR core loci are replacing, for the most part, all other current forensic DNA tests and have become the community standard for basic routine identity testing within the US. Several European laboratories are considering adopting these STR multiplex sets.

Several laboratories, particularly smaller units that started after 1994, never developed RFLP testing capability and instead began with simple PCR-based tests because of their smaller start-up costs and their relative technical simplicity. Many of these laboratories also correctly believed that DNA tests would gravitate toward PCR-based systems. PCR-based systems are quicker, less labor-intensive, more sensitive and applicable to degraded DNA specimens. The evolution of PCR-based tests has been a response to the need for small target DNA loci, the requirement for high-level discrimination of basic identity systems, the desire for discreteness of results, and the importance of high throughput by means of automation and robotics.

Future Genetic Systems

The recent achievement of a consensus core set of STR loci should standardize routine service casework and DNA databasing through 2005 and beyond within the US, and possibly globally. There had previously been poor consensus on DNA typing loci

and DNA typing systems. Thus, there are not likely to be great advances in standard identity genetic systems in the near future. Instead, future genetic systems will probably be developed to augment the 13 core STR loci, although there is some potential for new instrument platforms that will drive new systems, particularly to single nucleotide polymorphisms.

Mitochondrial DNA

Mitochondrial DNA (mtDNA) testing has been used for linking children to their grandmothers after generational exterminations (the so-called 'disappeared') in South and Central America, and for skeletal remains identification since the late 1980s. Since 1996, the FBI and others have used mtDNA testing for the forensic analysis of hairs.

Mitochondrial DNA may be successful when nuclear DNA is not, either because the tissue has little or no nuclear DNA, e.g. hairs, or because the nuclear DNA has become degraded and mtDNA is still available, owing to its initial high copy number, e.g. old skeletal remains. MtDNA is also useful when only limited kindred are available for reference specimens, e.g. when only a brother is available, due to an absence of generational dilutional effect because of a lack of crossover events and exact sequence transmission. The potential requirement for mtDNA testing in forensic laboratories is significant.

MtDNA analysis is currently performed using DNA sequencing by a relatively few laboratories. Most forensic laboratories now have the capability to perform DNA sequencing because they have instrumentation that can be used for sequencing. However, most laboratories are unaccustomed and unprepared to perform mtDNA sequencing. MtDNA sequencing is demanding and expensive due to poor template and the stringent requirements for preventing contamination. Several forensic laboratories are developing mtDNA sequencing capability and this number is expected to increase.

Relatively poor discriminatory power is a current limitation of mtDNA testing. It can be anticipated that mtDNA testing will expand beyond the areas of the hypervariable control region (D-loop) currently tested. Other regions, although less polymorphic, will add to the discriminatory power of mtDNA tests.

Another significant aspect of mtDNA sequencing is the potential for heteroplasmy, in which the same individual harbors more than one mtDNA sequence. Heteroplasmy is currently detected by the presence of a second peak in the sequencing electropherogram. This peak must be sufficiently above background noise for recognition. Newer, more sensitive, methods of heteroplasmy detection, such as denaturing gradient gel electrophoresis (DGGE), may prove to be of value.

The earliest mtDNA studies were performed by dot bolts. Dot blot analysis is less expensive and less technically demanding than sequencing. Efforts to commercialize mtDNA dot blots have not yet come to fruition, but may in the future. Several other technologies may be applied to mtDNA analysis, such as mass spectrometry, microchip instrumentation and molecular beacon analysis. Any of these technologies have the potential to make mtDNA analysis a more widespread and common procedure throughout the forensic DNA typing community.

The sensitivity of mtDNA analysis is far beyond other PCR-based techniques, due to the hundreds to thousands of copies of mtDNA per cell. It can be anticipated that contamination will cause difficulties for some naive laboratories during the implementation of this technology.

Y-chromosome markers

There has been an increasing interest, particularly in Europe, in Y-chromosome markers, not for gender determination but for identity. Y-chromosome markers may be useful for the typing of male DNA in the presence of contaminant female DNA, such as may be seen from the typical vaginal swab in sexual assault cases. This would not be necessary if current differential extraction procedures were efficient and complete, but some DNA is lost in the process and some female contaminant DNA is carried over. Y-chromosome markers might also be useful for tracing paternal lineages, just as mtDNA can be used to match maternal lineages. Some have argued that Y-chromosome markers will not be of great use in forensic laboratories, as other more polymorphic markers exist. Since the markers are linked, allele frequencies cannot simply be multiplied together. The number of Y-chromosome markers has increased and when used together may permit reasonable discrimination. This author views Y-chromosome markers as useful adjuncts; however, it might be possible in the future to truly automate vaginal swab DNA analysis through the use of such markers.

Alu repeats and other investigative markers

The commercial STR core loci kits include amelogenin, which is not an STR system but a gender marker. This marks a departure from all other routine identity markers as it provides specific phenotypic biological information about the source individual. Certainly, information about the gender of the DNA specimen may be an important investigatory lead in a criminal case. Amelogenin may also

be used simply to categorize specimens as originating from a victim or suspect. Amelogenin is useful as a gender determinant because the locus is a different size in X- and Y-chromosomes. Thus, two bands on a gel or electropherogram indicate a male, and a single band indicates a female. Other gender markers have been recognized and may be used to corroborate the amelogenin results.

Alu repeat markers are being developed to use as race and ethnicity investigation leads. The *Alu* families of short interspersed nuclear elements (SINEs) are distributed throughout the primate lineage and are the predominant SINE within the human genome. *Alu* elements are dimeric sequences of approximately 300 bp. The *Alu* family has spread throughout the genome by an RNA-mediated transposition process known as 'retroposition' and is present in the genome in extremely high copy number (in excess of 500 000 copies per haploid human genome). The majority of *Alu* family members are pseudogene products of a single 'master' gene. Sequence divergence in the master gene and its progeny occurs with time, resulting in subfamilies. Young *Alu* subfamilies are polymorphic and are present or absent on given chromosomes. Deletions and gene conversions are rare; thus the first appearance of the *Alu* insertion represents the beginning of the family tree and can be used as a molecular clock to estimate the time that family or subfamily arose. VNTR loci, including STR loci, are less stable and arise and vanish more frequently, resulting in ambiguity in ancestry from multiple convergent allelic states in time spans beyond a generation or two. Thus, unlike other forensic DNA markers, the distribution of *Alu* insertions, and possibly long interspersed nuclear elements (LINEs) and other SINEs loci, permit tracing of population ancestral heritages. This, in turn, permits a higher level of investigatory information to be garnered from an evidence sample. Information about the likely ethnicity of the sources of the sample is one piece of information that investigators may use when pursuing leads based on the genetic analysis of crime scene evidence.

Interest may become renewed in genetic systems which have previously been recognized for their disparate population frequencies, such as the Duffy blood typing system. Phenotypic markers for height, eye color, etc., may also become relevant for investigative purposes.

Investigative markers will generally not be conclusive, but rather suggestive. They may, in fact, hinder an investigation where there is a low probability occurrence. As they are not incriminating, it is not likely that they would be used as evidence in court.

Single nucleotide polymorphisms

Genome centers and clinical laboratories have increasingly looked at single nucleotide polymorphisms (SNPs) as markers of choice because SNPs represent the ultimate in the trend toward smaller DNA target fragments. SNPs are more numerous than other polymorphisms. They occur in coding and noncoding regions throughout the genome. SNP testing can be polymerase-based, ligase-based or probe hybridization-based assays. Very clean unambiguous results can be obtained easily. Band stutter is not problematic. The very small target regions are optimal for use on degraded specimens. The most important attribute of SNPs is their suitability to new automated instrumentation platforms, particularly mass spectrometry and microchip instrumentation, as well as in-solution techniques such as molecular beacon and fluorescence polarization. However, many SNPs need to be typed to achieve suitable discriminatory power. This is problematic in that some markers may reside within the same chromosome and thus their allele frequencies may not be able to be multiplied together, i.e. reducing discriminatory power. Given the emphasis on SNPs by molecular biologists outside the forensic arena, it can be assumed that interesting applications of these loci will develop.

Species determination and nonhuman individual and cluster identification

To date, the forensic typing community has largely been interested in the individual or so-called 'private' polymorphisms of humans. Nonhuman animal and plant DNA typing is becoming more common. Nonhuman polymorphisms may be private or group, subspecies, species, genus, family or class specific. Molecular systematics is the discipline that focuses on the phylogenetic relationship of such polymorphisms.

The US Fish and Wildlife Laboratory makes forensic use of species and subspecies tests to identify the booty of poachers. Forensic entomologists have employed species determination to assist their efforts to date postmortem intervals from the sarcophagous insects found on decomposing bodies. Determination of a stain or bone as human may also be performed using genetic markers. Species determination is usually performed using probes against relevant markers in conserved coding regions, such as ribosomal DNA.

The few cases where DNA typing of nonhuman biological samples have been of use in criminal trials to date have involved identification of an individual identification rather than of the determination of the species of origin. These cases have been unique, with little widespread application (e.g. Snowball the cat

and a Paloverde tree in Arizona); however, the potential for widespread application is great. Pet hairs have been transferred from suspects to scenes and vice versa. One can imagine that plant subspecies determination or identification might be very useful for marijuana tracing. One can also imagine that grasses found on the shoes of suspects might be important and common evidentiary specimens to link suspects to crime scenes. Typical analyses for nonhuman identification have included RAPDS, A-flps, ribosomal DNA and mitochondrial cytochrome B studies. Adequate validation may be problematic in the wide variety of applications entailed.

Future Instrumentation

The first widespread instrumentation in the forensic DNA community, other than basic equipment, were computerized image readers for RFLP autoradiographs. Current instruments use automated gel scanners (e.g. Hitachi FMBio, Molecular Dynamics FluorImager, BioRad MultiImager) and real-time gel detection systems, including slab gel electrophoresis (e.g. PE Biosystems 377) and capillary electrophoresis systems (e.g. PE Biosystems 310). New instrument platforms are available that offer the ability for far greater throughput than current instrumentation.

The new instruments should be characterized as batch, on-demand or field devices. Each type of instrument can produce the same DNA typing result but is suited for specific laboratory niches. Batch instruments analyze many samples at once and can achieve the highest specimen throughput. Such instruments are highly efficient when used to capacity. Batch testing is most appropriate for databasing operations where large numbers of uniform specimens are constantly being run.

On-demand instruments are ideal for casework. Forensic analysts generally work criminal cases one at a time. Batch instruments are inefficient when they are not used to capacity. The number of specimens to be tested in a case or set of cases may not precisely correspond to the number of specimens batched together for running on a batch instrument. High throughput is achieved by: (1) the ability to run the test when the analyst is ready to run the test; and (2) by fast turnaround time per test. Throughput is measured by the number of cases worked per analyst and not by how many specimens can be analyzed on an instrument in a given time.

Lastly, field instruments are rugged, portable, simple instruments that permit preliminary testing at the scene. Mass spectrometers hold the greatest potential as batch instruments and microchip technologies hold great promise as on-demand and field instruments.

Multichannel capillary electrophoresis

The fastest run times using instrumentation currently in service in forensic laboratories are achieved by capillary electrophoresis systems. These instruments are limited by their incorporation of only a single capillary and throughput is based on serial sample feed. Multichannel instruments are now available (PE Biosystems 3700, Molecular Dynamics MegaBACE 1000) which achieve huge throughput by parallel sample feed, e.g. 96 channels. These instruments are very expensive and their use does not make sense except in large databasing operations.

Mass spectrometry

Matrix-assisted laser desorption time-of-flight mass spectrometry (MALDI-TOF MS) is an exciting technology that has the capability of performing genetic analysis within seconds. With robotic sample preparation and sample loading, literally thousands of analyses have been performed in a single day. Furthermore, the precision of the sizing measurements is beyond any current technology. Extremely small specimens are analyzed. The operational cost per test is very low; however, the instrumentation is quite costly and trained operators are essential.

MALDI-TOF MS involves a solid-state specimen in a matrix which, when energized by a laser, imparts an ionic charge and ejects the analyte into the vacuum of the mass spectrometer. Since the kinetic energy is set by the laser wavelength, the mass can be calculated from the speed at which the analyte travels down the flight tube. Fragments are sized by their atomic weight. The precision will permit the detection of an altered nucleotide base sequence and has been used to sequence DNA of small size.

Unfortunately, MALDI-TOF MS is possible only on small fragments. STRs can be performed by using truncated PCR shoulder regions. SNPs are perhaps even better suited to this instrument. The evolution of this technology is such that it may be expected that ever larger fragments will be analyzed. Gene Trace and Perkin Elmer, as well as several national laboratories are pursuing mass spectrometric DNA analysis.

Microchips

The same microchip technology that revolutionized the personal computer is now being applied to microfluidic systems to generate microelectromechanical systems (MEMS) devices and the creation of 'labs on a chip'. Manufacturing costs are very low when such chips are mass produced. Miniaturization permits massive parallelism, integrated electronics with local feedback mechanisms, and in some cases enhanced performance characteristics. The basic

molecular biological techniques of electrophoresis, thermal cycling and hybridization can be accomplished in microchip formats. Prototype instruments have already amply demonstrated potential capabilities and some have begun to be commercialized.

Microchannel capillary electrophoresis is an obvious further miniaturization of current capillary electrophoresis. Surprisingly, only very small channel lengths are necessary for detectable separations of DNA fragments. STR determinations have been reliably performed in millisecond electrophoretic runs. Multiple channels can be placed on a single microchip. PCR and capillary electrophoresis can be integrated on a single chip. PE Biosystems and Caliper as well as several universities, national laboratories and the Whitehead Institute are pursuing microchip-based capillary electrophoresis. Forensic applications are important to several of these efforts.

Arrays of thousands of oligonucleotide probes on microchips can be used for sequencing specimen DNA by a process known as sequencing-by-hybridization (SBH). Small oligomers of every possible combination can be synthesized on chips through a photolithographic process. Hybridization is detected by means of a fluorescent confocal microscope. Computer recognition of the specific and nonspecific hybridization pixel patterns permits interpretation of the sequence information. Affymetrix, Hyseq and others are pursuing SBH strategies.

Electronic stringency control is a process in which specimen DNA affinity for specific probes is analyzed by varying the charge of the probe pad. Thus, bioelectronic chip platforms can be built to detect specific genetic systems. Microchips are built with longer but fewer probes than typical SBH chips. Rapid analyses, on the order of minutes, are possible. Nanogen is the proponent for this technology and they have identified forensics as an area of interest.

Thermal cycling can be achieved by etching a reaction chamber in the silicon wafer and coating the surface with a resistive film that acts as a heater when a current is passed. The exquisite temperature control achieved can permit very rapid thermal cycling – on the order of 20 s per cycle. The tight temperature control also decreases nonspecific amplification. Thermal cycling on a chip will allow a PCR as a first step after specimen preparation, followed by electrophoresis or hybridization on the chip for detection. The silicon medium will permit interrogation of the reaction chamber by a light source and photodiode. The TaqMan assay is a method of kinetic PCR analysis in which a fluorescent species is produced for each amplicon by polymerase exonuclease cleavage of a quenching moiety. Molecular beacon technology is an alternative strategy involving hairpin probes that normally hold a fluorophore adjacent to a quencher until the probe hybridizes to a target sequence permitting fluorescent detection. Using these techniques the PCR product may be measured directly during PCR amplification. PE Biosystems, Cepheid and other companies, as well as several national laboratories, are pursuing microchip PCR technology.

Several other technologies are being investigated with microchips as appropriate platforms. Semiconductor nanocrystals are being investigated because an array of colors can be produced from a single wavelength source. Several electronic detection systems are being investigated. One such technique is being commercialized by Clinical Micro Sensors.

Sequencing

Sequencing is employed for mtDNA typing but not for other forensic DNA systems. Sequencing is becoming far more rapid, less expensive and more facile, creating the potential for full sequencing by service laboratories. A single microchip has been engineered to sequence the full human mtDNA genome. Some have argued that a full sequence is the gold standard for a DNA type. Since most polymorphic information is dispersed, nonsequence technologies are more efficient. Indeed, any single locus, although highly polymorphic, will not approach the discriminatory potential of multiple independent sites whose probabilities can be multiplied together. On the other hand, a single microchip can be used to sequence multiple sites and a mass spectrometer may likewise eventually sequence multiple sites sufficiently quickly for casework. Already microchip probe arrays exist for sequencing genes of whole organisms.

Field DNA Testing

Owing to inherent miniaturization, microchip instrumentation has allowed the production of portable units that can be used in the field. Disposable chips are possible. Specimen processing may be incorporated into the instrumentation and the units may be made simple and expedient for field use. A field device has achieved results within 7 min. Specimen preparative techniques are being incorporated into many of the devices being developed (see below).

Field DNA testing can be useful for screening subjects, particularly to rule out suspects, and permitting investigating agencies to focus in quickly on their suspects. DNA samples may be obtained from simple oral swabs. Field DNA testing may also be useful in quick determinations of evidence sorting, e.g. 'this bloodstain is the same as that bloodstain' or 'this bloodstain is from the victim.' It may be useful in the

interpretation of blood spatter pattern analysis. Perhaps of greatest value is the potential to perform an immediate search against existing databases of open cases and convicted offenders.

It is this author's belief that any field testing of an incriminating nature should be confirmed in the more optimal setting of a laboratory, ideally by a second technology. It is also the belief of this author that field DNA tests should remain in the hands of laboratory personnel who travel to the field, rather than being conducted by nonscientific police personnel.

Some have argued that there is no need for field testing, particularly laboratory personnel, but many scene investigators have articulated their desire for such a capability. Prototype technology exists now and commercialization will make it available, rendering adoption inevitable.

Some have argued that field testing will abrogate the rights of the citizenry. The fourth amendment (prohibition against search and seizure without probable cause) has been used as the basis for this argument. Others argue that taking a DNA sample for forensic identity investigations is no different than obtaining fingerprints.

Apart from forensic identity testing, field DNA testing is relevant to bioterrorist events where a biological agent must be identified. Several recent hoaxes have involved the sending of a white powder labeled 'anthrax' to government and business facilities. Entire city blocks have been shut down while the deposited material is sent out to determine the nature of the material. It is now clear that these are issues of local law enforcement. A field DNA testing device would greatly facilitate the handling of these situations. Public jeopardy demands an immediate response.

Trace DNA Testing

Recent investigations have indicated that trace DNA can be obtained from fingerprints. DNA types have been obtained from a surprising number of substrates after handling. Such evidential trace specimens are probably commonplace. DNA testing is particularly likely to be of benefit where the fingerprint is smudged. Clean fingerprints are often difficult to find on bodies and paper, although they can sometimes be obtained. Weapons, particularly handguns, represent the most obvious type of evidence where fingerprints are rarely found, where the need to link the suspect is great, and where transferred DNA is likely. Another significant reason for obtaining a DNA profile is for purposes of CODIS database searching.

Casework testing must be optimized for greater sensitivity to detect such minute traces. Trace DNA studies to date have involved nuclear DNA, but the more sensitive mtDNA testing may have greater application.

Trace DNA testing will be problematic. The potential for contamination is an obvious issue for such testing. Investigations into the longevity of trace DNA in various environments and on various substrates will be important.

Sample Preparation and Characterization

Relatively little attention has been paid to sample preparation, clean-up and characterization. Differential extraction procedures to separate the female from the male DNA of sexual assault swabs could be improved and automated. Recently a two-swab method, a wet swab followed by a dry swab, for collecting DNA from bite marks on bodies has been described. Such optimization of collection techniques should continue for other substrates. The utility of sonication, microwave and detergents in processing some types of specimens has been demonstrated but not pursued within the forensic arena. The release of DNA by *N*-phenacylthiazolium from the protective effect of sugar–protein condensate (thiazole) formation, useful in recovering fossil DNA from coprolites, deserves investigation. The 'ancient DNA' community also uses whole genome amplification and other techniques to recover minute quantities of damaged DNA.

There is a strong commercial incentive to develop methods of processing specimens as part of integrated microchips. Several approaches have been described.

Characterization of the quantity and quality of DNA in a specimen may also be improved. Current techniques were developed a number of years ago and new technologies permit better and more rapid analyses. Slot blot quantitation is a relatively insensitive, labor intensive and somewhat imprecise quantitation method. Competitive PCR assays and quantitative TaqMan and molecular beacon assays may become useful replacements. DNA damage has not been assessed in forensic laboratories, but may be important in the case of low-copy, environmentally challenged specimens as a potential source of error.

PCR Amplification Alternatives

Patents and commercial realities have driven many to search for alternatives to PCR amplification. PCR has proven to be a more robust technique than many of its initial competitors. Interest continues in several systems, such as the ligase chain reaction and strand displacement amplification. The looming expiration of the PCR patent has quenched the interest of some.

On the other hand, the desire to create and commercialize microchip instruments that perform all relevant steps for obtaining a DNA profile has renewed interest of others in the use of alternative systems. New systems have been introduced, such as that by Third Wave Technologies based on a Cleavase 'invader' system. Most of these alternative techniques are isothermal and thus better adapted to automated and microchip systems.

Robotic Sample Preparation

High throughput analytic instruments are now available, and even higher throughput machines are on the near horizon. Most laboratories are now overwhelmed with databasing and casework, yet the demand for DNA testing is ever increasing. Such testing capacity outstrips most laboratories' ability to feed these machines and creates an impetus to automate specimen preparation and specimen loading. Many robotic systems are available and others are being developed to meet this need. Initial demand for such systems have originally come from genome project centers and pharmaceutical companies that require a capacity beyond that of most forensic laboratories, hence the ironic need to scale back many current robotic systems. At least one commercial effort (Rosys) is aimed squarely at forensics; it would begin with stacks of DNA cards and generate PCR products ready for loading on to a slab or capillary electrophoretic systems.

Computer Analysis and Laboratory Information Systems

The forensic community is reaching a point where the time to review, interpret and report DNA test results is taking longer than the actual laboratory testing itself. Thus, computers linked to laboratory equipment that automates interpretation and reporting will relieve a bottleneck in current or near future efforts. Determination of base pair calling confidence for sequencing operations provides an excellent example of the potential of computerization enhancements. Computer analysis can provide input to an analyst as to which bases of a sequence are in doubt, and thus direct the analyst's time in a focused, efficient manner, as well as ensuring that a questionable base call is not overlooked. Computers may perform other review and quality assurance functions for virtually any type of analysis. Reporting turnaround times and reduction of clerical errors are significant achievements of laboratory information management systems, in addition to evidence and case tracking functions. Advances in such systems will become an emphasis for the community.

Conclusion

Although we may have reached a momentary plateau with respect to current identity testing through the acquisition of the current generation of instrumentation and the adoption of a standard core set of STR markers, there are many areas in which forensic DNA typing efforts will continue to advance rapidly. We will see advances in adjunctive utilization of new genetic systems and technologies. But beyond the current plateau, next generation technologies will fundamentally change the nature of forensic DNA typing and dramatically increase the role of forensic science in law enforcement. Only next generation technologies can bring DNA testing to the field, and only next generation technologies are capable of meeting the future potential demand of forensic DNA testing. Instruments will more closely resemble 'black boxes'. Forensic DNA analysts will process far greater case loads. Interactions with law enforcement agencies will decrease. Laboratory work will increasingly resemble nonforensic clinical and industrial production laboratories.

See also: **Analytical Techniques**: Mass Spectrometry. **Deoxyribonucleic Acid**: Restriction Fragment Length Polymorphism; Polymerase Chain Reaction; Polymerase Chain Reaction-Short Tandem Repeats; Significance; Databanks; Mitochondrial Deoxyribonucleic Acid.

Further Reading

Belgrader P, Bennett W, Hadley SD, Richards J, Stratton P, Mariella R and Milanovich F (1999) PCR detection of bacteria in seven minutes. *Science* **284**:449–450.

Belgrader P, Smith JK, Weedn VW and Northrup MA (1998) Rapid PCR for identity testing using a battery-powered miniature thermal cycler. *J Forensic Sciences* **42**:315–319.

Butler JM (2000) *Forensic DNA Typing: Biology and Technology Behind STR Markers*. London: Academic Press

Butler JM, Li J, Shaler TA, Monforte JA and Becker CH (1998) Reliable genotyping of short tandem repeat loci without an allelic ladder using time-of-flight mass spectrometry. *Int J Legal Med* **112**:45–49.

Hodgson J (1998) Shrinking DNA diagnostics to fill the markets of the future. *Nature Biotechnology* **16**:725–727.

Holland MM and Parsons TJ (1999) Mitochondrial DNA sequence analysis – validation and use for forensic casework. *Forensic Science Review* **11**(**1**): 21–50.

http://www.cepheid.com/

http://www.cstl.nist.gov/div831/strbase
http://www.microsensor.com/core.htm
http://www.nanogen.com/
http://www.ojp.usdoj.gov/nij/dna/welcome.html
Kricka LJ (1998) Revolution on a square centimeter. *Nature Biotechnology* **16**:513–514.
Sensabaugh G and Kaye DH (1999) *Jurimetrics*, **39**: 1–16.
Schmalzing D, Adourian A, Koutny L, Ziaugra L, Matsudaira P and Ehrlich D (1998) DNA sequencing on microfabricated electrophoretic devices. *Analytical Chemistry* **70**(**11**): 2303–2310.
Sinha G (1999) DNA Detectives. *Popular Science Magazine* 48–52.
Siniscalo Gigliano G (in press) *Cannabis sativa* L: botanical problems and molecular approaches in forensic investigation. *Forensic Science Review*.
Stoneking M, Fontius JJ, Clifford S, *et al.* (1997) Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. *Genome Research* 7:1061–1071.
Van Oorschot RAH and Jones MK (1997) DNA fingerprints from fingerprints. *Nature* **387**:767.
Wang DG, Fan JB, Siao CJ *et al.* (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280**:1077–1082.
Weedn VW and Hicks JW (1997) The unrealized potential of DNA testing. *NIJ Journal* **234**:6–23.
Wilson EK (1998) Instant DNA detection. *Chemical and Engineering News* **76**:47–49.

Mitochondrial Deoxyribonucleic Acid

T Melton, Mitotyping Technologies and Department of Anthropology, Penn State University, State College, PA, USA
G Sensabaugh, School of Public Health, University of California, Berkeley, CA, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0776

Introduction

DNA profiling is routinely used to link individuals to biological evidence found at scenes, on victims or on suspects. In most cases, DNA profiling involves analysis of highly discriminating genetic markers located on the chromosomal DNA found in the cell nucleus. However, in some situations, e.g. the analysis of aged or burned skeletal remains, hair fragments, and degraded stains, it is not possible to extract enough usable nuclear DNA from the evidence sample to do the profiling analysis. In these situations, it may be possible to use an alternative approach: genetic analysis of the DNA present in cell mitochondria. The genetic analysis of mitochondrial DNA is also a preferred approach when the forensic question entails establishing identity through familial relationship, e.g. the identification of recovered remains from the war dead or victims of mass disasters. Because mitochondrial DNA is directly inherited through the maternal lineage, it is possible to make familial connections over multiple generations provided a maternal lineage can be clearly traced.

Mitochondria and Mitochondrial DNA

Mitochondria: biology and inheritance

Mitochondria are small organelles contained within almost all animal and plant cells; red cells are the only major cell type that do not have mitochondria. Mitochondria function as the energy powerhouses of the cell; they transmute the energy released by the metabolic breakdown of sugars, fats and amino acids into high-energy molecules (e.g. ATP) that drive most of the metabolic processes within the cell. Mammalian cells typically contain several hundred to several thousand mitochondria per cell.

The DNA in mitochondria (mtDNA) is distinct from the DNA in the cell nucleus in five important respects. First, the mtDNA genome is very small compared to the nuclear genome. The human mtDNA genome, which has been completely sequenced, is a circular piece of DNA, ca. 16 569 bp around; it encodes 37 genes. The nuclear genome, in contrast, is linear, about 3 billion bp long, and encodes an estimated 100 000 genes. Second, in humans and in most if not all mammals, mtDNA is maternally inherited; that is, mtDNA is passed from mother to offspring without change and there is no evidence of a paternal contribution. (Although there are mitochondria in sperm, these are sequestered and destroyed when the sperm enters the oocyte.) Nuclear DNA, in contrast, is inherited equally from both mother and father in accord with Mendelian genetics. Third, as a consequence of uniparental inheritance, each individual typically possesses a single mtDNA genetic type that can be unambiguously determined by DNA sequence analysis. With nuclear DNA, the genetic type reflects the sequence contributions of both mother and father. Fourth, because cells contain many mitochondria (each with 2–10 copies of mtDNA) but only one nucleus, there are hundreds to thousands more copies of mtDNA than nuclear DNA per cell. Last, mtDNA

appears to accumulate mutations more rapidly than nuclear DNA; thus there is more variation per unit length of DNA sequence in mtDNA than in the nuclear genome. These features contribute to the attractiveness of using mtDNA in forensic testing, particularly in situations where the amount of DNA in an evidence sample is limiting or the question to be answered involves a familial relationship along the maternal lineage.

mtDNA variation

The structure of the mtDNA genome is illustrated in **Fig. 1**; nucleotide positions in the sequence are numbered 1–16 569 according to a standard reference sequence, the Anderson (or Cambridge) reference sequence. Interindividual variation in human mtDNA is concentrated in the control region, so named because it contains several sequence elements that control replication of the mtDNA molecule. (The control region is often referred to in the older literature as the D-loop region.) The control region extends from nucleotide position 16 024 through the position 16 569/1 boundary on to position 576 (Fig. 1). Within the control region, there are two hypervariable regions, HV1 (16 024–16 365) and HV2 (73–340); these two regions are the primary targets for analysis in forensic identity testing.

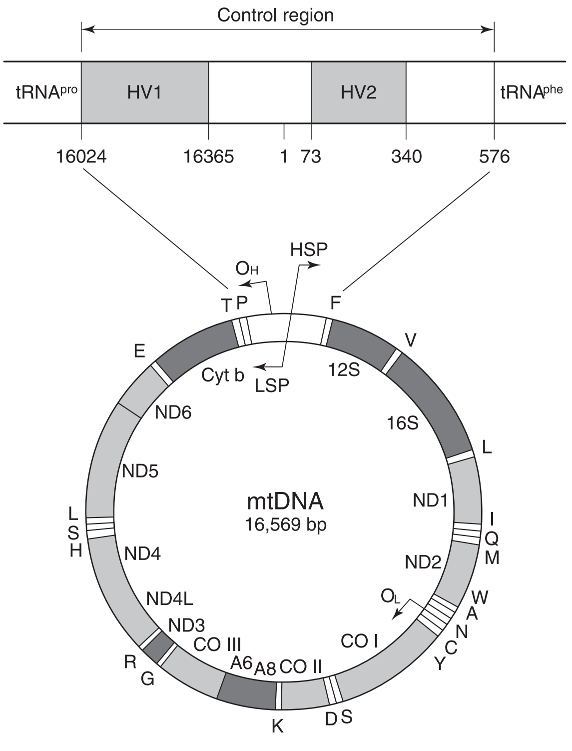


Figure 1 Map of the human mitochondrial genome and expanded diagram of the noncoding control region. Gene designations (inside ring) are 12S (12S ribosomal RNA), 16S (16S ribosomal RNA), ND (NADH coenzyme Q complex subunits), CO (cytochrome oxidase complex subunits), Cyt b (cytochrome *b*), and A (ATP synthase). The 22 tRNA genes are indicated around the outside of the ring in the single letter amino acid code. The arrows indicate light strand and heavy strand origins of replication (O_L and O_H) and light strand and heavy strand transcriptional promoters (LSP and HSP). The numbering on the control region diagram refers to the Anderson reference sequence. (Reproduced with permission from Holland and Parsons (1999) *Forensic Science Review* **11:**21–50).

The extent of interindividual variation in the mtDNA control region is graphically illustrated by pairwise comparison of sequences from unrelated individuals. **Figure 2** shows the results of pairwise comparison of sequences from 100 unrelated individuals of the same racial group (British Caucasians); the average sequence difference between any two individuals in this sample set is 8.5 with a range of 0–25 differences. The chance that two individuals from this sample set do not differ in their control region sequences is about 0.004. Pairwise comparison data within other racial and ethnic population groups show the same pattern of variation (**Table 1**). When individuals from different population groups are compared pairwise (Table 1), the average number of sequence differences tends to increase and the frequency of a chance sequence match tends to decrease; this shows simply that separated populations tend to accumulate different patterns of nucleotide substitutions. When a very large number of pairwise sequence comparisons are done using individuals without regard to population group, the frequency of a chance match is less than 0.002 (Table 1).

The low frequency of a chance match in the data in Table 1 reflects the fact that most individuals have different mtDNA control region sequences. However, some sequence types are seen in more than one individual and a small number of sequence types

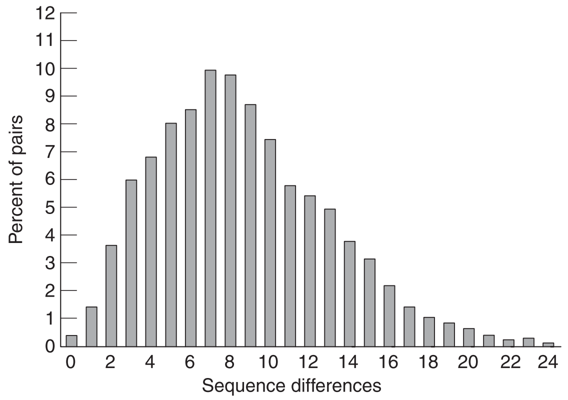


Figure 2 Pairwise comparisons in mtDNA control region sequences from 100 individuals. (Redrawn from data in Piercy R, Sullivan KM, Berson N and Gill P (1993). The application of mitochondrial DNA typing to the study of white caucasion genetic identification. *International Journal of Legal Medicine* **106:** 85–90.)

Table 1 Pairwise comparison of mtDNA control region sequence types within and between populations

Population (pairs compared)	*Frequency of matches*	*No. of differences in sequence*	
		Average	*Range*
Within population			
African (6105)	0.0099	13.1	0–31
French (5886)	0.0039	7.2	0–19
Hispanic (4851)	0.0074	12.6	0–30
Japanese (13041)	0.0040	9.7	0–23
Between populations			
African/French (12099)	None seen	13.3	3–28
African/Hispanic (10989)	None seen	15.3	1–33
African/Japanese (17982)	None seen	13.8	2–30
French/Hispanic (10791)	0.0005	11.7	0–27
French/Japanese (17568)	None seen	9.9	2–22
Hispanic/Japanese (16038)	None seen	12.1	2–27
Total population			
Eight populations (969528)	0.0016	10.6	0–36

can be regarded as almost common. This point is illustrated using the population data characterized in Table 1 representing control region sequence analysis on 1393 individuals drawn from different racial and ethnic groups. In this population, 1063 distinct sequence types were detected of which 932 were seen in single individuals, 19 were present in five or more individuals and the most common sequence type was found in 38 individuals. These observations underscore the importance of having a large database to use as the basis for interpreting the results of any particular sequence typing.

Sequence variation in the mtDNA control region has been extensively investigated as part of studies on human population history. Indeed, the fact that the greatest amount of sequence variation has been observed among African populations is primary evidence for the African origin of modern humans. Control region sequence data have both provided support for and challenged previously held notions of relationships between populations. These data have also given new insights into human migration patterns, e.g. the waves of migration across the Bering Strait during the last Ice Age. It is sometimes possible to characterize individuals as belonging to a particular population group based on that individual having a particular control region sequence motif.

Heteroplasmy

Most individuals have a single mtDNA control region sequence type. However, on occasion an individual has been found to carry two sequences that differ at one or two nucleotide positions; this is referred to as heteroplasmy. Heteroplasmy may take two forms. Site heteroplasmy occurs when two different nucleotides are detected at a single nucleotide position. Length heteroplasmy occurs in C nucleotide runs when different numbers of Cs are detected in the run; there is a C nucleotide run in both HV1 and HV2. The extent of heteroplasmy in humans is difficult to assess since the detection of multiple sequences depends on the analytical sensitivity of the sequencing technology used; the most sensitive technologies detect heteroplasmy in up to 15% of individuals but conventional sequencing methods detect heteroplasmy in only 1–2% of individuals. Heteroplasmy tends to be found at nucleotide positions that are also frequent sites of variation between individuals; this is consistent with the idea that the same mutational mechanisms are responsible for heteroplasmy and population variation.

In comparing control region sequences of maternally related individuals, two patterns of heteroplasmy are seen: (1) the individuals may show the same pattern of heteroplasmy or (2) some of the individuals are heteroplasmic and others appear homoplasmic for one or the other of the heteroplasmic sequences. In the former case, the fact of shared heteroplasmy supports the commonality of lineage. The latter case makes it clear that single nucleotide sequence differences do not necessarily indicate that individuals belong to different maternal lineages.

Heteroplasmic individuals may exhibit different proportions of the heteroplasmic sequences in different tissues. In most tissues, the differences are recognizable and pose no problem for forensic testing. With hair, however, there may be an uneven distribution of

heteroplasmic sequences in different hairs resulting in a situation where a particular hair may exhibit one or the other or both of the heteroplasmic sequences. Thus when the evidence to be tested is hair, it is important to evaluate the possibility of heteroplasmy.

Technology of Detection

Sequencing analysis

The typical mtDNA sequencing analysis is a four-step process. First, DNA is extracted from the biological sample such as bone, blood or hair. Current extraction methods can capture DNA from as little as 0.25 g of bone or 0.5–2 cm of a hair fragment. Second, the HV1 and HV2 regions are amplified as multiple overlapping fragments using the polymerase chain reaction (PCR). These fragments serve as the templates for sequence analysis; the overlaps allow multiple sequencing of each of the sequence regions. The third step is the sequencing reaction using the Sanger dideoxy nucleotide approach. The sequencing reaction products are differentially labeled with fluorescent dyes such that each of the four nucleotides, A, T, C and G, is represented by a different label. The last step is the reading of the sequence on an automated sequence analyzer. The sequence thus determined is read into a computer where it is analyzed, assembled into the two hypervariable region sequences, and compared to the published standard reference sequence, the Anderson reference sequence. Typically, there is 3–4 fold redundancy in the sequence determination of the HV1 and HV2 sequence regions; this provides confidence in the obtained sequences.

Contamination control is very important in forensic mtDNA analysis. Because many of the samples that are selected for mtDNA analysis contain minimal amounts of DNA to begin with and because PCR-based methods of analysis are extremely sensitive, introduction of only a few strands of contaminating DNA or PCR product into a sample tube or reagent or onto an item of equipment might result in amplification of the contaminant rather than the sample of interest. Accordingly, much care is taken to avoid contamination during both the extraction and amplification steps. Contamination control measures include use of prepackaged sterile equipment and reagents; use of aerosol-resistant barrier pipette tips, gloves, masks and lab coats; separation of pre- and postamplification areas in the lab; use of dedicated reagents in each area; ultraviolet irradiation of equipment and autoclaving of tubes and reagent stocks. It is good practice to process the questioned samples separately from the reference samples to which they are to be compared; this can be done by processing the questioned samples, typically containing minimal DNA, first, or, if lab facilities allow, by processing questioned and reference samples in separate laboratory rooms. An additional contamination control measure is the use of negative control samples which are processed through the four steps of analysis in parallel with the test samples; if there is contaminating DNA in the reagents or on equipment, it will be detected in the negative control samples.

Interpretation of Results

Comparison of sequences is facilitated by first identifying the sites in each sequence that differ from the Anderson reference sequence. This simplifies the comparison task to ‘substitutions with respect to Anderson’; the numbered sites where the two samples differ from each other are easily seen in a table format. Sequences are compared according to guidelines based on the known biology of mtDNA and an established body of experience with mtDNA sequencing. Sequence comparison can have several possible outcomes.

- Sequences that differ at two or more nucleotide positions are scored as an exclusion; they do not originate from the same person or from persons in the same maternal lineage.
- Sequences that are identical at every sequence position are scored as an inclusion; they either originate from the same person, from persons in the same maternal lineage, or from persons who happen by chance to share the same mtDNA sequence.
- Sequences that differ at a single site with evidence of heteroplasmy at that site are also scored as an inclusion. Differences in the number of Cs in either of the two control region C nucleotide runs are not considered determinative.
- Sequences that differ at a single site with no evidence of heteroplasmy are regarded as nondeterminative, i.e. neither a match nor a nonmatch; the comparison is scored as inconclusive. The results are considered inconclusive because the possibility that a complete mutational change occurred between mother and offspring without a transitory heteroplasmic state cannot be excluded; the possibility of a mutational change between the mtDNA in different tissues (blood and hair, for example) or of an undetected heteroplasmy also cannot be excluded.

If the sequence comparison shows an inclusion, the significance of the inclusion must be estimated. As

noted above, an inclusion indicates that the sequences either originate from the same person, from persons in the same maternal lineage, or from persons who happen by chance to share the same mtDNA sequence. The probability of detecting any particular sequence by chance can be estimated from databases of mtDNA control region sequences. A search of the database can provide information about the number of observations of any particular sequence, overall and by racial group. A 95% or 99% confidence interval (for previously observed types) or upper bound (for novel types) may also be calculated around the projected true frequency of a sequence type in the general population, given the current size of the database. As discussed previously, about two-thirds of the sequences in the databases are unique; the continual addition of new sequences to the databases is not changing this proportion.

Other methods of detecting variation in mtDNA

Sequencing the two hypervariable regions is the most comprehensive and definitive approach to forensic mtDNA analysis. However, there are situations in which a laboratory might want to make a quick assessment of the overall variation in a population or might need to screen many samples to determine which samples are possible inclusions and qualify for sequencing. A number of procedures have been developed for the rapid detection of differences between sequences; all, however, are capable of detecting only a portion of sequence variants. Methods applied to mtDNA analysis include the following.

- Sequence specific hybridization probes can be used to test amplified mtDNA control region sequences; this is typically done in a dot–blot or reverse dot–blot format.
- Single strand conformation polymorphism (SSCP) analysis and denaturing gradient gel electrophoresis (DGGE) typically allow differentiation of about 70% of the sequence variants.
- Heteroduplex analysis entails hybridizing complementary strands of mtDNA from different samples; sequence differences result in strand mismatches that can be detected by electrophoresis.
- Allele specific PCR uses PCR primers that bind to alternative sequences at a variable site; an amplification product will be generated only if the primer finds its sequence matched site.
- Minisequencing detects single nucleotide variation at highly variable sequence positions.

Although these approaches allow rapid screening of samples for sequence differences, they carry the disadvantage that a portion of the extracted mtDNA is consumed on a nondefinitive test. Moreover, the risk of contamination increases every time a sample is opened. As DNA sequencing technology becomes quicker, easier and cheaper, there will be less need for alternative tests.

Other Applications of mtDNA Testing

The identification of the species origin is sometimes a matter of forensic interest, particularly in forensic wildlife investigations and in the analysis of foodstuffs. Two mtDNA genes, the cytochrome *b* gene and the cytochrome oxidase I gene, have been extensively investigated in evolutionary studies and the sequence data from these studies provide the basis for forensic applications. The process of species identification usually entails sequence determination of the target gene. If the obtained sequence matches a sequence in the database, the species is identified. If the obtained sequence matches no sequence in the database, phylogenetic analysis is used to determine the most closely related species.

See also: **Deoxyribonucleic Acid**: Restriction Fragment Length Polymorphism; Future Analytical Techniques; Parentage Testing; Significance; Databanks.

Further Reading

Anderson S, Bankier AT, Barrell BG *et al.* (1981) Sequence and organization of the human mitochondrial genome. *Nature* **290**:457–465.

Budowle B, Wilson MR, DiZinno JA *et al.* (1999) Mitochondrial DNA regions HVI and HVII population data. *Forensic Science International*:**103**:23–35.

Handt O, Meyer S and von Haeseler A (1998) Compilation of human mtDNA control region sequences. *Nucleic Acids Research* **26**:126–129.

Holland MM and Parsons TJ (1999) Mitochondrial DNA sequence analysis – validation and use for forensic casework. *Forensic Science Review* **11**:21–50.

Kogelnik AM, Lott MT, Brown MD, Navathe SB and Wallace DC (1998) MITOMAP: a human mitochondrial genome database – 1998 update. *Nucleic Acids Research* **26**:112–115.

Lightowlers RN, Chinnery PF, Turnbull DM and Howell N (1997) Mammalian mitochondrial genetics: heredity, heteroplasmy and disease. *Trends in Genetics* **13**:450–455.

Meyer R and Candrian U (1996) PCR based DNA analysis for the identification and characterization of food components. *Food Science Technology* **29**:1–9.

Stoneking M (1996) Mitochondrial DNA variation and human evolution. In: Jackson M, Strachan T and Dover G (eds) *Human Genome Evolution*. Oxford: BIOS Scientific Publishers.

Parentage Testing

M S Schanfield, Analytical Genetic Testing Center, Denver, CO, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0472

Introduction

The need to identify the parents of children is as old as history. A famous biblical example is that of two women coming to King Solomon and each arguing that a living infant was hers, after one of the women's child had died. Solomon's solution was to have a sword produced and threaten to cut the child in two, thus providing each mother with a half. One mother agreed to let the other have the child; the other mother agreed to let the child be killed. Solomon decided the 'true' mother to be the one who chose to give up the child rather than have it killed, and awarded the child to her.

Parentage identification has three forms: paternity identification, maternity identification and parentage identification. Paternity identification is widely used for purposes of providing financial support to children in Europe, Canada and the United States. Paternity identification can also be important in cases of inheritance. From the forensic standpoint, paternity identification can be important in cases of sexual assault in which a pregnancy occurs and the pregnancy is either terminated or goes to term; the identity of parentage of the fetus or newborn child is evidence of a sexual act. Similarly, in child abandonment cases, the establishment of maternity is evidence implicating the alleged mother in either abandonment or infanticide. Finally, a critical use of parentage identification involves missing body homicides, in which crime scene evidence can be genetically profiled and used to establish whether the missing individual is the parent or child of known individuals.

Modern parentage identification began with the rediscovery of mendelian genetics and the discovery of polymorphic variation in humans. The discovery of blood group(s) antigens on the red blood cells (ABO, Rh, MNS, Kell, Duffy, Kidd) allowed for the exclusion of approximately 50% of falsely accused fathers in cases of disputed paternity. The addition of the immunoglobulin allotypes (GM and KM), and plasma protein markers identified by electrophoresis (haptoglobin (HP), group-specific component (vitamin D binding globulin) (GC/VDBG), transferrin (TF)) and red cell enzymes (acid phosphatase (ACP/EAP), phosphoglucomutase (PGM1), glyoxalase (GLO1), esterase D (ESD) etc.) all increased the ability to exclude falsely accused fathers. Later, the addition of the highly informative markers on white blood cells (HLA), and additional variation in plasma proteins and red blood cell enzymes through the use of isoelectric focusing, not only allowed for the high likelihood of excluding falsely accused parents but, in civil cases, allowed for the calculation of likelihood and probability of paternity. The advent of DNA technology, both in the form of restriction fragment length polymorphisms (RFLPs) to detect length variation caused by the presence of variable number tandem repeats (VNTRs), and later polymerase chain reaction (PCR)-based technology in the form of DQA1 dot blots or amplified fragment length polymorphisms (AFLPs) detecting VNTRs of different lengths (long tandem repeats (LTRs) and short tandem repeats (STRs)), increased the ability to determine parentage. At the present time the majority of paternity testing is being done using STRs. With the advent of DNA-based testing, non-DNA testing has steadily declined, such that in the United States in 1997 only approximately 6% of cases were routinely tested for red blood cell antigens, enzymes or protein markers. In contrast, approximately 45% of cases were routinely tested for PCR-based loci, and approximately 40% of cases were routinely tested for RFLP loci.

DNA Testing in Disputed Paternity

Paternity testing (parentage testing involving an alleged father) is highly organized and regulated in both Europe and the United States because of the pressures of individuals applying for child support. Almost 240 000 paternity determinations were done in the United States in 1997. Parentage testing has a long history in the United States and Europe, such that guidelines have been developed for the testing of parentage cases for conventional markers (red cell antigens, plasma proteins and red cell enzymes). In the last several years, two organizations (the International Society of Forensic Hemogenetics (ISFH) and the American Association of Blood Banks (AABB)) have developed guidelines for DNA RFLP- and PCR-based testing in cases of disputed paternity. Many of these guidelines are similar to those developed by the FBI advisory panel (Technical Working Group on DNA Analysis Methods (TWGDAM)); however, some features unique to parentage identity testing are included. Paternity testing has several advantages over forensic identity testing. The primary advantage is the quality and quantity of sample. Normally, for RFLP testing 1–10 μg, per lane of DNA to be tested, are digested with a restriction enzyme. The AABB standards require that a mixture of the alleged father's and child's DNA are run in the same lane to

verify that bands are in fact the same. This is true whether the test is an RFLP or STR. Chemiluminescent nonisotopic detection for RFLP-based DNA testing has increased the usefulness of the technology by eliminating the need for radioisotopes, while keeping sensitivity and decreasing the amount of time needed to perform the testing.

Genetics of Parentage Identification

Parentage identification is exclusionary, as is all forensic identification. The object of the testing is to eliminate as large a portion of the population as possible. The ability to exclude an individual is determined by Mendel's laws of inheritance and not by a comparison to a phenotype array as in forensic identification. Mendel's laws are presented in **Table 1**.

Mendelian Laws of Parentage Testing

The first two mendelian laws have been restated to create four rules for parentage testing. These rules are presented in **Table 2**. When the questioned parent is the father, it is assumed that the mother is the biological mother. However, when the questioned parent is the mother, no assumptions are made about the mother or the father.

If exclusions violate rule 1 or 2, they are called 'direct' exclusions because direct evidence indicates that the alleged parent cannot be the parent. Exclusions that violate rules 3 and 4 are called 'indirect' exclusions, because the exclusion is inferred from genetic information. It is inferred that the individual has two alleles that are the same. However, it would not be an exclusion if the 'single band pattern individual' had undetected alleles. In the case of RFLP technology, if bands run off the gel or are too large to be detected, then an individual will appear to have a single band and be excluded, when in fact this is not the case. These null alleles have a frequency of approximately 0.0004 for RFLP loci. For PCR-based systems, a mutation in a primer site can lead

Table 1 Mendel's laws of inheritence.

1. Crosses between individuals homozygous for two different alleles at one genetic locus produce individuals heterozygous for the two different alleles (**Fig. 1**)
2. When these heterozygous individuals mate with each other, different genotypes segregate out. This is referred to as the Law of Segregation (**Fig. 2**)
3. When individuals differing at more than one gene pair mate, every gene pair segregates independently, if the genes are unlinked, and the resulting segregation ratios follow the statistical law of independent assortment. This is referred to as the Law of Independent Assortment (**Fig. 3**)

From Vogel F and Motulsky A (1979) *Human Genetics, Problems and Approaches*. New York: Springer.

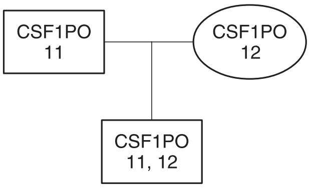


Figure 1 Mating of two parents homozygous (both alleles the same) for alleles 11 and 12 at locus CSF1PO. All children produced are heterozygous (both alleles different) for alleles 11 and 12. This is a demonstration of Mendel's first law.

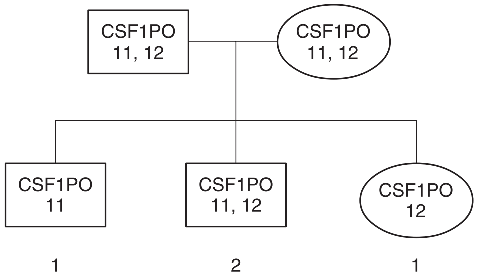


Figure 2 Demonstration of Mendel's second law. Mating of two parents heterozygous for alleles 11 and 12 at locus CSF1PO produce offspring that are homozygous for CSF1PO 11, heterozygous CSF1PO 11,12 and homozygous CSF1PO 12 in the ratio of 1:2:1.

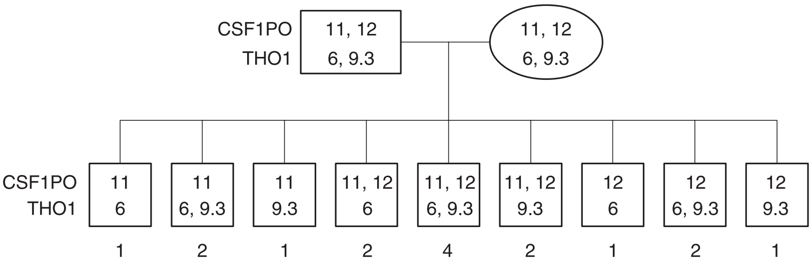


Figure 3 Demonstration of Mendel's third law (independent assortment). Mating of two parents heterozygous for two independent inherited traits: CSF1PO 11,12 and THO1 6,9.3. The expected ratio of the different types is represented under each child.

Table 2 Paternity rules based on Mendel's laws.

1. A child cannot have a genetic marker which is absent in both parents (**Table 3**)
2. A child must inherit one of a pair of genetic markers from each parent (**Table 4**)
3. A child cannot have a pair of identical genetic markers unless both parents have the marker (**Table 5**)
4. A child must have the genetic marker which is present as an identical pair in one parent (**Table 6**)

From Lee CL (1975) *Family Law Quarterly* **9:** 615–633.

Table 3 Examples of an inclusion and exclusion following paternity rule 1 (Table 2)

	Inclusion	*Exclusion*
Mother	VWA 16,**17**	VWA 16,**17**
Child	VWA ***15***,**17**	VWA **17**,***18***
Alleged father	VWA ***15***,17	VWA 15,**16**

Alleles inherited from the mother are bold printed; the biological father's alleles are bold, italicized and underlined.

Table 4 Examples of an inclusion and exclusion following paternity rule 2 (Table 2)

	Inclusion	*Exclusion*
Mother	VWA 16,**17**	VWA **16,17**
Child	VWA ***16,17***	VWA ***16,17***
Alleged father	VWA ***16***,18	VWA 15,18

Alleles inherited from the mother are bold printed; the biological father's alleles are bold, italicized and underlined.

Table 5 Examples of an inclusion and exclusion following paternity rule 3 (Table 2)

	Inclusion	*Exclusion*
Mother	VWA 16,**17**	VWA 16,**17**
Child	VWA ***17***	VWA ***17***
Alleged father	VWA ***17***,18	VWA 15,18

Alleles inherited from the mother are bold printed; the biological father's alleles are bold, italicized and underlined.

Table 6 Examples of an inclusion and exclusion following paternity rule 4 (Table 2)

	Inclusion	*Exclusion*
Mother	VWA **18**	VWA **18**
Child	VWA ***17***,**18**	VWA ***17***,**18**
Alleged father	VWA ***17***	VWA 16

Alleles inherited from the mother are bold printed; the biological father's alleles are bold, italicized and underlined.

to the failure to amplify one allele, creating a silent allele. The frequency of null alleles in PCR-based systems appears to be significantly higher than that for RFLP loci, at approximately 0.0023. In Table 5, the apparent exclusion could in fact be an inclusion, if the alleged father and child share a very small fragment that is not detected, or an inherited mutation in a primer site. Therefore, because of the possibility that an indirect exclusion is really an inclusion owing to an uncommon event, indirect exclusions are not considered to be as conclusive as direct exclusions. For this reason, two indirect exclusions or an indirect and direct exclusion are required before an exclusion is established.

The factors that make DNA polymorphisms so useful, i.e. the high degree of genetic heterogeneity, also can create problems. The genetic heterogeneity is caused by higher frequencies of mutation. The mutation rate for DNA that codes for proteins is reported to be 10^{-5}. In contrast, the average mutation rates for RFLP loci used in parentage identity testing is approximately 1.7×10^{-3}, or two orders of magnitude higher. Mutation rates for STR AFLP systems and LTR AFLP loci D1S80 and D7S5 appear to be similar to those for RFLP loci.

Some RFLP loci have reported mutation rates as high as 5% (DIS7/MSI). These loci are not normally used in parentage identity testing. Owing to the higher mutation rate at VNTR loci, whether detected by RFLP or PCR at least two exclusions are required by the organizations that regulate parentage testing (AABB Parentage Testing Committee for the United States, and the ISFH for Europe and the rest of the world).

Reference Populations and Databases

A question that has created a great deal of furor in forensic identity testing in the United States has to do with reference populations. In the United States the reference population for civil cases of disputed parentage is that of the questioned parent. In contrast, in forensic cases the population frequencies are calculated from all possibly relevant populations and the most conservative one is used. In some areas of the world this will be quite limited, but in the United States there may be several potential populations of donors. Issues such as substructuring and the Hardy–Weinberg equilibrium (HWE) are not relevant to parentage identity testing because they affect genotype frequencies and not allele frequencies. Since in parentage testing it is the frequency of people in the population that can transmit the obligate allele, and not the likelihood of getting a specific genotype. Therefore, any population databases which have

been validated for forensic testing can be used for parentage identity testing, assuming they are the appropriate reference populations.

Mathematics of Forensic Parentage Identification

If the evidenciary material cannot be excluded, then it is useful to know what portion of the population could have donated the material, or could be the parent of the material in question. Obviously, the larger the battery of tests the higher the likelihood of excluding someone. The current standards of the AABB require that the battery of tests have an average exclusion power of 99%, while in Germany, it is said that the average power of exclusion must be 99.99%. Ideally, the average power of exclusion should be high, so as not to falsely include someone. The ability to exclude individuals is determined by Mendel's laws. The frequency of the markers in potential parents is determined by the behavior of genes in the populations. Genes behave following the binomial distribution in populations. Although the formulas look identical to the HWE, the rules governing the behavior of genes is the binomial. The HWE applies to the behavior of genotypes in the population.

Parentage testing primarily deals with determination of paternity. The determination of maternity, sibship, grandparentage and other forms make up a small number of cases usually treated as variations, as are paternity without the mother, or 'motherless' cases. All are special cases. In the United States there are major differences in the mathematics of paternity depending on whether the case is criminal or civil. These differences do not necessarily occur outside of the United States. In the United States and elsewhere, the likelihood ratio or paternity index and the bayesian probability of paternity (hereafter referred to as the probability of paternity) are used, whereas in criminal cases, the population frequency of nonexcluded potential fathers (also referred to as the random man not excluded or RMNE method) is used. Bayes theorem is compatible with other population frequencies used in forensic science to individualize evidence. The two sets of calculations are based on very different perspectives. The paternity index (likelihood ratio) and probability of paternity are based on calculations using the genetic type of the nonexcluded alleged father to calculate the specific likelihood or chance that this alleged father is the biological father. In contrast, the RMNE method is based on the child, which is comparable to the evidence, and calculating from the information known about the child's genes that had to come from the biological father. The RMNE method asks what portion of the population could be the father. The probability of paternity is not admissible in criminal cases in the United States because it assumes a prior probability that the alleged father is the biological father. In most cases this is 0.5. The courts have equated this with a prior assumption of guilt, and it has not been allowed under the United States constitutional protections. These same protections do not exist elsewhere. Likelihood ratios and bayesian probability are used in England and possibly elsewhere in criminal cases.

There are additional problems with the likelihood ratio/probability of paternity when compared to the RMNE or nonexclusion probability. Under some circumstances, when using AFLP (LTR or STR), blood groups, protein or enzyme genetic markers, more testing does not yield an increase in either the paternity index or probability of paternity. In fact it can decrease. In contrast, the RMNE/nonexclusion probability method always increases as a function of the number of tests performed. As the paternity index can be converted to a probability of paternity, the RMNE can be used to calculate a post-test bayesian nonexclusion probability of paternity or more precisely the probability that the alleged father was 'truly accused' based on the mendelian inheritence. The bayesian probability based on the nonexclusion probability is conveniently 1 minus nonexclusion probability, when a prior probability of 0.5 is used. In the United States a prior probability of 0.5, often called a neutral prior probability, is routinely used. Since there are many ways of determining the prior probability, it was to some extent arbitrarily decided in the United States and elsewhere to use the prior probability of 0.5, with the option of using others. The author has not seen a request for other prior probabilities in many years. Therefore, unless needed for comparative purposes, only the equivalent of the population frequency will be presented, as this requires significantly fewer formulas than the parentage index calculations and are readily comparable.

Paternity Identification

Table 7 contains a comparison of 'forensic identity testing' with 'paternity identity testing'. In forensic identity testing, if the evidence tested is not excluded, the frequency of the event is calculated from the phenotype frequency. Thus, using STR locus CSF1PO, with the type 10, 11, the frequency would be $2pq$, where p is the frequency of the *CSF1PO*10* allele, and q is the frequency of the *CSF1PO*11* allele. For blood groups, enzymes and protein markers the calculations are the same as for the AFLP (LTRs and STRs), as these systems use

Table 7 Comparison of forensic identity testing with forensic paternity testing using the RMNE

Identity testing	*Paternity testing*
Phenotype *CSF1PO 10,11*	Obligate ailele *CSF1PO1*10*
Frequency $2pq$	$2p(1-p)+p^2$
p = *10 = 0.264	p = *10 = 0.264
q = *11 = 0.267	$1-p$ = 0.736
$2pq$ = 0.141	$2p(1-p)+p^2$ = 0.458

defined alleles. For RFLP-based systems the calculations are the same; however, since they are based on fragment sizes, the frequency is determined by either a fixed or a floating binning window.

For parentage testing, when the mother is available for testing, it is assumed that the mother is the biological mother, and the obligate allele or alleles (the allele(s) that could not have come from the mother) are identified based on mendelian inheritence. In **Table 8** the child has the alleles *CSF1PO*10* and *CSF1PO*11*; as the mother can only contribute *CSF1PO*11* or *CSF1PO*12*, the child had to receive the *CSF1PO*10* from its biological father. The *CSF1PO*10* allele is referred to as the obligate allele in this case. All potential fathers without *CSF1PO*10* would be excluded as possible fathers, while all potential fathers with *CSF1PO*10* could not be excluded. This includes potential fathers with the two *CSF1PO*10* alleles (phenotype = CSF1PO 10) and all potential fathers with one *CSF1PO*10* allele and another allele (CSF1PO 10, 10′ where the 10′ indicates all alleles other than 10). This converts to the formula

$$p^2 + 2p(1-p) \qquad \textbf{(Equation 1)}$$

where p is the frequency of the allele *CSF1PO*10*. If the mother and child have the same type, i.e. share two alleles, there are two possible obligate alleles. The formula when the mother and child have the same genotype, reduces to

$$(p+q)^2 + 2^*(p+q)^*(1-[p+q]) \qquad \textbf{(Equation 2)}$$

where p and q represent the two different alleles. This population frequency of the possible fathers is referred to as the random man not excluded (RMNE). Note: tested potential fathers are referred to as either the ‘alleged father’ or the ‘putative father’ in civil paternity testing in the United States, but usually as the suspect in criminal cases.

For US European, African, Mexicans and Asians, the frequency of the *CSF1PO*10* allele is respectively 0.264, 0.255, 0.291 and 0.278, and the RMNEs are 0.458, 0.445, 0.497 and 0.479. Due to the larger numbers of possible parents at each locus tested, the combined RMNE is not as individualizing as the results seen in forensic identity testing.

Routine Paternity Identity Testing

This is the most common type of paternity testing worldwide and is normally performed at the request of civil authorities to obtain child support. It can occur in criminal cases due to incest or rape if the pregnancy could not be terminated, or if the crime was not discovered until late in the pregnancy or after the birth of the child. Paternity identity testing on liveborn children in the United States is usually done either on whole blood or buccal swabs. Civil paternity testing as regulated by the AABB requires the reporting of the paternity index and probability of paternity. In the United States forensic paternity testing uses the calculations described above in Equation 1 or 2.

Prenatal Paternity Testing

Prenatal paternity testing is possible when either amniocentesis or chorionic villus sampling (CVS) is performed for medical or legal reasons. Medical reasons for prenatal paternity testing include a fetal risk of inheriting a genetic disorder for which no direct test exists, but a possible parent is known to have a genetic disorder. Women over the age of 35 are at increased risk of having children with chromosomal abnormalities such as Down syndrome. Therefore, prenatal

Table 8 Example of calculation of RMNE in a sample forensic parentage case

Mother	CSF1PO **11**,**12**			
Child	CSF1PO ***10***,11			
Alleged Father	CSF1PO 9,***10***			
Calculation of RMNE	$= p^2 + 2p(1-p)$ (1)			
White example	$= (0.264)^2 + 2^*0.264^*0.736$			
	= 0.0697 + 0.3886			
	= 0.4583			
	RMNE in several populations			
	European	*Black*	*Mexicans*	*Asian*
*CSF1PO*10*	0.264	0.255	0.291	0.278
RMNE	0.4583	0.4450	0.4973	0.4787

screening is routinely performed on pregnant females over the age of 35. Therefore, if a question about the parentage of the child exists, prenatal paternity testing can be performed. The majority of prenatal paternity tests performed in our laboratory have been this type of situation. Legal reasons for prenatal paternity testing are that the pregnant female was sexually assaulted while in a consensual sexual relationship. If the fetus originated from the rapist the victim normally terminates the pregnancy. The results of the test can then be used as evidence against the suspected perpetrator. Other legal reasons include suspected cases of incest, or that the mother was mentally incapable of giving consent. The latter case usually occurs when a mentally retarded individual is sexually assaulted while in a custodial environment.

On an ethical side, there appears to be general consensus in the United States that prenatal paternity testing will not be performed if there is neither a medical nor legal reason for obtaining the necessary fetal samples.

Amniocentesis

Amniocentesis is the older of the two procedures. A sample of amniotic fluid is normally obtained about the sixteenth week of gestational age (with a range of 12–20 weeks) by inserting a needle into the amniotic sack, removing amniotic fluid containing fetal cells, and growing up the fetal cells using sterile tissue culture techniques. This is performed routinely in laboratories doing clinical genetic testing. When cases of this type are referred for testing, the referring laboratory is responsible for growing the needed cells. When using RFLP DNA technology it is necessary to obtain 2–4 tissue culture flasks with 25 cm^2 of growing surface, referred to as T25 flasks. Yields will vary, but about 20 μg of HMW DNA can be obtained from four T25 flasks (range 6.5–44 μg). This is more than enough DNA to perform RFLP testing. With the shift to PCR-based STR testing, less DNA is needed. Although we still perform the tests from DNA obtained from T25 flasks, recent experiments indicate that it may be possible to extract DNA from the cells obtained from amniocentesis without the need for growing the cells first.

Chorionic villus sampling

Chorionic villus sampling (CVS) is normally performed between 8 and 12 weeks' gestation. It involves the collection of placental tissue and processing the tissue to make sure that little or no maternal tissue contaminates the fetal tissue. There is a great deal of variation in the ability of physicians to collect 'clean' CVS. Due to difficulties in obtaining nonmaternally contaminated tissue, it is necessary to test the DNA from the CVS and the mother, to verify that the CVS DNA is fetal and not either predominantly maternal or a mixture.

How much DNA can be obtained from CVS? Thus far, the average is approximately 10 μg (range 7.5–12.5 μg) of HMW DNA per CVS.

The testing of CVS tissue by DNA technology has been previously reported by one group on testing four CVS samples ranging from 7 to 22 weeks, and another reported on the testing of a single CVS specimen, using PCR-based LTR AFLP loci.

Products of Conception Paternity Identity Testing

Products of conception (POC) is the term applied to material obtained after the termination of a pregnancy, either naturally (stillbirth or miscarriage) or induced. In the United States, this covers material obtained as early as 7–8 gestational weeks or as late as 24 gestational weeks. POC cases are almost always criminal in nature, usually arising from the sexual assault of a minor, with the perpetrator a blood relation (incest) or not (carnal knowledge/sexual assault), or from the sexual assault of an adult when the pregnancy is known to be associated with the sexual assault. Since paternity testing is not normally performed by public crime laboratories in the United States, cases of this type are normally referred to commercial laboratories. However, in Europe forensic laboratories also often carry out paternity testing. Several laboratories have reported on the results of testing that they have performed.

The amount of tissue and the development of the fetus will affect what source material is used to obtain genomic DNA. A 6–8 week gestational age POC, if obtained from an induced termination in the USA, is usually produced by a suction abortion. The fetus is very small and hard to identify, except for occasional organs, making extraction of DNA from fetal tissues difficult. On the other hand, the 6–8 week POC consists of large amounts of chorionic villus material, which is easy to tease out and identify, yielding large amounts of HMW DNA. Fetuses in the 10 week plus range normally have readily identifiable organs for obtaining DNA, so that either single organs such as a kidney or muscle can be readily extracted. With the exception of a single case, in which the fetus had been stored for 1 year at refrigeration temperatures, ample DNA is usually available for either RFLP- or PCR-based DNA technology. The biggest difficulty in testing fetal tissue occurs if the material has been placed in formalin to preserve it. This normally only

happens if the POC is obtained at autopsy. It is still possible to obtain results with PCR-based testing, but it can be limited.

RFLP technology is a powerful tool in eliminating falsely accused suspects in cases of parental identity testing. It is normally possible to eliminate 99.9% of falsely accused suspects with as few as 3–4 probes. For comparative purposes, three RFLP POC cases are presented in **Table 9**, to illustrate the exclusionary power observed with the same four probes in European-American, African-American and Southwestern Hispanic-American POC cases. The loci have the same approximate level of heterozygosity in each population but the most informative locus differs in each case. This points out that there is no way of knowing which system will provide the most information. In all cases the combined exclusionary power (1-RMNE) is greater than 99.9%. That is to say that the likelihood of excluding a falsely accused suspect exceeds 99.9%. With the introduction of PCR-based STR AFLP testing, many cases previously tested by RFLP will be tested by PCR. STR AFLP loci are not as polymorphic as RFLP loci, so more loci are needed to reach the same level of discrimination. **Table 10** gives comparative data from three consecutive STR AFLP POC cases from three US populations, using the eight locus Powerplex 1.1 test from Promega.

Effect of Related Suspects

If there are multiple suspects and they are first-degree relatives, the exclusionary power drops significantly. **Table 11** illustrates the case of the stepfather of a minor who was charged with her sexual assault when the victim was discovered to be pregnant. An initial analysis of DNA from the victim, stepfather and the fetus, using three probes, did not exclude the stepfather as the father of the fetus. However, at trial, it emerged that there was a 1 in 8 chance that two stepbrothers (teenage sons of the stepfather) could not be eliminated as possible fathers. (Note: The likelihood that first-degree relatives, such as siblings, would share any given allele is $\frac{1}{2}$; therefore, the likelihood that they would share three different regions is $\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}$, or $\frac{1}{8}$. The judge ordered three laboratories to perform additional testing. Due to the nature of the material, one laboratory opted to perform PCR-based typing, while the other two laboratories performed RFLP testing. Stepbrother 1 was excluded by multiple RFLP systems but stepbrother 2 was only excluded by a single RFLP system. In contrast, the PCR-based LTR AFLP systems did not exclude stepbrother 1, but had two exclusions of step-brother 2. The stepfather was not excluded by any of the systems tested.

The lesson in this case is that there is no way of predicting which systems will provide information in these cases, and that when dealing with first-degree relatives it can be very difficult to differentiate them. It is therefore necessary to make sure that adequate numbers of genetic marker systems have been tested to differentiate first-degree relatives. As a rule of thumb, 10–12 highly polymorphic markers would be needed. In the case cited above, a total of 17 DNA loci were tested to obtain three exclusions on stepbrother 2. The likelihood that 14 loci would match by chance is 1 in 16 000.

Comparison of Paternity Identity to Forensic Identity Testing

As stated above, prenatal, POC or liveborn forensic paternity identity testing yields results significantly less individualizing than those observed in forensic identity testing. This is due to the fact that in paternity identity it is the RMNE that is used as the unit of comparison, versus the phenotype frequency in forensic identity testing. To provide a direct comparison, the three cases presented earlier as RFLP POC cases were recalculated as forensic identity cases, using the phenotype of the suspect father. **Table 12** presents the

Table 9 Population frequencies observed in POC cases from different populations using RFLP loci

Population	US White	US Black	US SW Hispanic
Case no.	93-011201	91-032602	91-010803
Locus (probe)			
D1S47(pL336)	**0.107**	0.207	0.201
D7S104(pS194)	0.372	0.230	**0.026**
D17S34(p144-D6)	0.378	**0.020**	0.567
D21S112(pL427-4)	**0.034**	0.267	0.109
Combined RMNE	0.0005	0.0003	0.0003
Exclusionary power (1-RMNE) (%)	99.95	99.97	99.97
Likelihood of coincidental match	1/2000	1/3700	1/3000

All RFLP systems used the enzyme *PstI*.

Table 10 Comparison of RMNE using STR-AFLP loci

Population	European-American	African-American	SW Hispanic
Case no.	98-123001	98-111803	98-121701
Locus			
D16S539	0.2483	0.3632	**0.2221**
D7S820	0.6808	0.5788	**0.3916**
D13S317	**0.2758**	0.5225	0.3341
D5S818	0.5670	**0.5349**	0.6031
'GammaSTR'	0.0264	0.0587	0.0175
CSF1PO	0.3963	0.5335	**0.0988**
TPOX	**0.3374**	0.4627	0.7269
THOI	**0.2894**	0.6069	0.4043
VWA03	**0.3806**	**0.3806**	0.5639
CTTV	0.0147	0.0570	0.0164
Powerplex RMNE	0.0004	0.0034	0.0003
Exclusionary power (%)	99.96	99.66	99.97
Likelihood of coincidental match	1/2600	1/294	1/3500

Table 11 PCR based forensic paternity (POC) involving first-degree relatives (case no. 93-010508)

AFLP	*Mother*	*Fetus*	*Stepfather*	*Stepbrother 1*[a]	*Stepbrother 2*[b]
APOB	34,40	40,**48**	34,48	34,48	*34*
PAH	8,10	3,10	**3**,8	**3**,9	*8*
D1S80	26,31	**24**,**26**	**18**,**24**	**24**	18,**24**
D4S95[c]	1.03,1.07	**1.03**,**1.07**	**1.03**	**1.03**	**1.03**,1.15

[a] Step-Brother 1 was excluded by multiple RFLP probes
[b] Step-Brother 2 was only excluded by a single RFLP probe out of 13.
[c] LTR AFLP with 40 bp repeat, number of repeats unknown.
Bold-faced underlined alleles represent the obligate alleles. Italicized results indicate nonpaternity.

Table 12 Comparison of paternity identity versus forensic identity at four RFLP loci

Population	Case no.	Paternity	Identity[a]	Ratio
US White	93-011201	2000	460 000/6 600 000[b]	230/3300
US Black	91-032602	3700	14 000 000	3 800
US Mexican	91-010803	3000	1.9×10^9	636 000

[a] Calculated using a 2% floating bin, minimum number of bands in a bin is 5 to be compatible with FBI-type fixed binning.
[b] D17S34 has a single band pattern. If the FBI $2p$ is used the combined value is 460 000. If p^2 is used, as would be appropriate in this case, the combined value would be 6 600 000.
To keep the numbers comparable, no corrections for sample size were used.

combined RMNE for the three cases, along with the likelihood of a coincidental match (referred to as 'Identity' in the example). It is obvious that the 'Identity' results are orders of magnitude higher. In fact the 'Ratio' expresses how many more times the 'Identity' result is than the 'Paternity' result. The extreme divergence in the Mexican case occurred because the nonobligate allele was uncommon for two of the four loci. Similar calculations are performed in **Table 13** for eight STR loci, only this time the child's results were used. Again the ratio indicates an enormous difference between parentage identity and forensic identity. With STRs the ratio appears to be magnified.

Maternity Identification

The need to identify the mother can occur in cases of child abandonment or infanticide. In these cases nothing is usually known about the biological father. Occasionally, in the case of a missing body only a single parent is available, creating a similar situation. Therefore, if the child in question is heterozygous (has two different alleles), both of the alleles are obligate alleles. The formula for the frequency of potential mothers in the population or random female not excluded (RFNE) is the same as Equation 2. In **Table 14** the obligate alleles found in the child are *CSF1PO**10* and *CSF1PO**11*. The mother of the

Table 13 Comparison of Paternity Identity versus Forensic Identity at eight STR[a] loci

Population	Case no.	Paternity	Identity[b]	Ratio
US White	98-060303	1260	4.2×10^8	334 297
US Black	98-092402	552	3.3×10^9	6 054 651
US Mexican	98-110501	302	4.7×10^8	636 000

[a] All samples tested with Powerplex 1.1 (D16S539, D7S820, D13S317, D5S818, CSF1PO, TPOX, THO1, VWA).
[b] Calculated using NRC 2 guidelines, including a minimum allele calculation binning.

Table 14 Example of calculation of RFNE in a sample forensic maternity identity case

Mother	CSF1PO **11**, 12			
Child	CSF1PO **10**,**11**			
Calculation of RFNE	$= (p + q)^2 + 2^*(p + q)^* (1-[p + q])$			
White example	$= (0.264 + 0.267)^2 + 2^*(0.264 + 0.267)^*(1-0.264-0.267)$			
	= 0.282 + 0.498			
	= 0.780			
	White	*Black*	*Mexican*	*Asian*
CSF1PO*10	0.264	0.255	0.291	0.278
CSF1PO*11	0.267	0.234	0.289	0.243
RFNE	0.780	0.739	0.824	0.771

child must share either *CSF1PO*10* or *CSF1PO*11* or she is excluded. However, any female with either *CSF1PO*10* or *CSF1PO*11* could not be excluded as a possible mother. Using the same data tables as above, the RFNE values are 0.780, 0.739, 0.824 and 0.771 for European-Americans, African-Americans, Mexican-Americans and Asian-Americans, respectively. If the child only has a single allele, Equation 1 is used.

Identification of maternity using just the deceased infant and the suspect mother is not very informative, as shown above. Two examples involving infanticide are presented in **Table 15**, one from a European-American and one from an African-American, in which data was available for four RFLP DNA systems and for 10 non-DNA genetic markers (six plasma proteins: GM, KM, GC/VDBG, F13A, F13B and C6 and four red cell enzymes: ACP1, ESD, GLO1 and PGM1). The four DNA loci do not reach an RFNE of 1% (exclusionary power of 99%) in either of the two cases. If data from two more probes are added to case 92-060401, the RFNE reaches 1% (data not provided). The 10 non-DNA genetic markers provide RFNE values comparable to the four DNA loci. However, most of the information in the non-DNA markers is provided by GM and KM allotypes (RFNE 0.032 and 0.082 for the two cases, respectively). These markers provide unique information in cases

Table 15 Two cases of forensic maternity identification from infanticide cases

Population	European-American	African-American
Case no.	92-060401	90-060401
RFLP loci		
D1S47/pL336 *PstI*	0.394	0.426
D7S104/pS194 *PstI*	0.369	0.293
D18S17/pL159-2 *PstI*	0.549	0.486
D21S112/pL427-4 *PstI*	0.484	0.352
Combined	0.0386	0.0214
non-DNA[a]	0.0192	0.0672
Combined	0.0007	0.0012
Likelihood of coincidental match	1 / 1360	1 / 817
Exclusionary power (%)	99.93	99.88

[a] Non-DNA markers tested include: GM, KM, GC (VDBG), F13A, F13B, C6, PGM1, ACP1, ESD and GLO1.

of infanticide because the GM and KM markers in newborns are maternal in origin and therefore represent a blood stain left at the crime scene by the biological mother of the infant. The GM and KM phenotypes represent forensic identity rather than RFNE. Individualization at the 1% level is possible using only DNA markers but it usually takes more than four RFLP loci or 12 STR loci.

Missing Body Cases

The mathematics of missing body cases is slightly different but uses the same rules of segregation seen in mendelian genetics. In **Table 16** a blood stain is found at a crime scene and is believed to have originated from a missing victim. The results among other markers are CSF1PO 10,11. The parents of the victim are tested and cannot be excluded as the parents of the blood stain found at the crime scene. The question is, 'What portion of the population could not be excluded as the parents of the stain?' To be consistent with the other terms used, the author offers the term random parents not excluded (RPNE) In this case, there are two obligate alleles, *CSF1PO*10* and *CSF1PO*11*. They will each have to be contributed by one parent, and each will have the distribution in the population of $p^2 + 2^* p^* (1-p)$ and $q^2 + 2^* q^* (1-q)$, respectively. However, since the question being asked is what is the chance that two people at random would produce this individual, the individual terms are multiplied to obtain the frequency for obtaining a set of parents of the blood stain, or

$$(p^2 + 2p[1-p])^*(q^2 + 2q[1-q]) \quad (\textbf{Equation 3})$$

For a discussion of likelihood ratio tests and Bayesian probability the author would refer the readers to Evett and Weir, 1998.

The mathematics for one type of missing body case was presented above (Table 16). Sometimes, by testing parents, wives and children, it is possible to determine the exact genetic type of the missing person and compare that to the evidence. That information would allow for a forensic identification.

If an individual is missing under what appears to be mysterious circumstances, and there is crime scene evidence found, often blood or tissue, the question asked is whether this came from a given person. This type of case occurs in most jurisdictions infrequently. The second type of case involving the identification of remains discovered is becoming much too common. Tragedies in the form of aircraft crashes or explosions and mass graves during civil or other acts of aggression have left thousands of remains, found and unfound. Depending on the duration of time that has passed since the incident, environmental insult and the cause of death, remains may have enough intact DNA to test nuclear DNA by RFLP- or PCR-based testing; however, in most cases it appears that PCR is the best tool for the analysis of nuclear DNA markers. In many cases the DNA is too badly degraded and can only be tested using mitochondrial DNA. Since mitochondrial DNA is inherited only in the maternal line, the comparison is to the mother or maternally related individuals.

Table 17 presents an example of an identification of remains from the Balkan conflict. The case is from the original publication but the example was recalculated using data available from Croatia. The bayesian probability of parentage was calculated using a prior probability of 0.5. If additional STR markers had been available it is likely that the results would have been over 99.9%.

Summary and Conclusions

In summary, parentage identification testing has been used widely worldwide in civil courts for many years. Although the statistical treatment of forensic parentage testing only represents a subset of the calculations used in paternity testing, at least in the United

Table 16 Example of the calculation of RPNE

Mother	CSF1PO **11**, 12			
Blood stain	CSF1PO ***10***,11			
Father	CSF1PO 9,***10***			
Calculation of RPNE	$= (p^2 + 2p[1-p])\ ^*\ (q^2 + 2q[1-q])$ (Equation 3)			
White example	$= (0.264^2 + 2^*0.264^*[1-0.264])^*(0.267^2 + 2^*0.267^*[1-0.267])$			
	$= (0.0697 + 0.3886)\ ^*\ (0.0713 + 0.3914)$			
	$= (0.4583)\ ^*\ (0.4627)$			
	$= 0.2121$			
	White	*Black*	*Mexican*	*Asian*
RFNE	0.212	0.184	0.246	0.204

Frequencies from Table 7 used.

Table 17 Identification of bone remains using parents[a]

Sample	*Mother*	*Bone 4*	*Father*	*Allele*	*P*[b]	*Parent*	*RPNE*
DQA1	2,4	3,4	2,3	3	0.105	0.198	
				4	0.340	0.565	0.112
LDLR	B	A,B	A	A	0.410	0.652	
				B	0.590	0.832	0.542
GYPA	A,B	A,B	A,B	A	0.560	0.506	
				B	0.440	0.686	0.554
HBGG	A	A,B	B	A	0.530	0.779	
				B	0.470	0.719	0.560
D7S8	A	A,B	A,B	A	0.653	0.880	
				B	0.347	0.574	0.505
GC	A,C	A,C	A,B	A	0.279	0.480	
				C	0.595	0.836	0.401
				Combined population of possible parents			0.004
				Probability of parenthood			99.62%

[a] Taken from Primorac D *et al.* (1996) *Journal of Forensic Sciences* **41**: 891–894.
[b] Data recalculated from Northern and Southern Croatians to form pooled Croatians from Keys KM *et al.* (1996) *Forensic Science International* **81**: 191–199.

States, the principles and limitations are exactly the same.

The introduction of DNA technology, both RFLP- and PCR-based, allows for the generation of much higher levels of individualization, in most types of cases, and broadens the spectrum of types of specimens that can be analyzed. However, new hazards, such as increased mutation rates, are also introduced. The majority of these potential pitfalls have been addressed by guidelines developed by the AABB Committee on Parentage Testing and the ISFH for RFLP- and PCR-based parentage testing. Although DNA-based technologies are powerful tools, it should not be forgotten that the old, reliable non-DNA protein and enzyme markers can provide additional information that can enhance the results obtained by DNA technology.

Owing to differences in the nature of the population frequencies when doing parentage identity testing, and the difficulty in eliminating first-degree relatives as suspects, it is often necessary to use more than the five loci normally used in forensic RFLP identity testing to reach satisfactory levels of individualization.

Finally, due to the changing nature of the case load, at least in the United States, forensic scientists are now being faced increasingly with the identification of assailants in sexual assault cases, using fetal cells, tissue or blood, collected as prenatal, postnatal or postmortem material, the identification of 'deceased victims' who are not there, and the identity of parents of abandoned and murdered children.

See also: **Evidence**: Statistical Interpretation of Evidence/Bayesian Analysis. **Deoxyribonucleic Acid**: Statistical Analysis.

Further Reading

AABB Parentage Testing Standards Committee (1998) *Annual Report Summary for 1997*. Fairfax, VA: American Association of Blood Banks.

Andradas J, Garcia E, Cámara T, Prieto L and López J (1996) Identification of human remains using DNA amplification (PCR). *Advances in Forensic Haemogenetics* 6:258–260.

Arroya E, Garcia-Sanchez F, Gomez-Reino F, Ruiz de la Questa JM and Vicario JL (1994) Prenatal exclusion of paternity by PCR-FLP analysis of VNTR. *Journal of Forensic Science* 39:566–572.

Augustin C, Groβevollmer E and Püschel K (1998) Identification of human remains: short tandem repeat analysis versus mitochondrial DNA. *Progress in Forensic Genetics* 7:83–85.

Baum HJ, Fritz-Charles H and Mckee R (1990) A non-isotopic DNA detection system with the sensitivity of ^{32}P: application for paternity and forensic identifications. *Advances in Forensic Haemogenetics* 3:37–39.

Brandt-Casadevall C, Dimo-Simonin N and Krompecher T (1996) Identification through genetic typing of the victims of the Sect of the Solar Temple (Chiery/Salvan Switzerland). *Advances in Forensic Haemogenetics* 6:243–245.

Callen DF, Thompson AD, Shen Y *et al.* (1993) Incidence and origin of 'null' alleles in the (AC)n microsatellite markers. *American Journal of Human Genetics* 52:922–927.

Corach D, Sala A, Penacino G and Sotelo A (1996) Combined STR, VNTR and MVR typing and mtDNA sequencing, led to the identification of human remains emergent from the AMIA explosion in Buenos Aires. *Advances in Forensic Haemogenetics* 6:337–339.

D'Errico G, Zanon C, Lago G, Vecchio C and Garofano L (1998) DNA analysis on skeletal remains after months of permanence in sea-water. *Progress in Forensic Genetics* 7:106–108.

Dobosz M, Pascali VL, D'Aloja E, Pescarmona M and Fiori A (1990) Exclusion of an individual charged with rape by allele-specific DNA profiling on fetal tissue. *Advances in Forensic Haemogenetics* **3**:117–118.

Evett I and Weir B (1998) *Interpreting DNA Evidence*. Sunderland: Sinauer.

Giles AF, Booth KJ, Parker JR *et al.* (1990) Rapid, simple non-isotopic probing of Southern blots for DNA fingerprinting. *Advances in Forensic Haemogenetics* **3**:40–42.

Hammond HA, Jin L, Zhong Y, Caskey CT and Chakraborty R (1994) Evaluation of 13 short tandem loci for use in personal identification applications. *American Journal of Human Genetics* **55**:175–189.

Keys KM, Budowle B, Andelinovic S *et al.* (1996) Northern and southern Croatian population data on seven PCR-based loci. *Forensic Science International* **81**:91–199.

Lee CL (1975) Current status of paternity testing. *Family Law Quarterly* **9**:615–633.

Li CC and Chakravarti A (1988) An expository review of two methods of calculating the paternity probability. *American Journal of Human Genetics* **43**:197–205.

Liss LR and Hudson GR (1990) Comparison of chemiluminescent and radioactive methods of DNA typing. *Advances in Forensic Haemogenetics* **3**:48–50.

Livovksy A and Schultz K (1998) Scientific evidence of paternity: a survey of state statutes. *Jurimetrics Journal* **39**:79–94.

Lobbiani A, Nocco A, Vedrietti P, Brambati B and Colucci G (1991) Prenatal paternity testing by DNA analysis. *Prenatal Diagnosis* **11**:343–346.

Ludes BP, Mangin PD, Malicier DJ, Chalumeau AN and Chaumont AJ (1990) Parentage determination on aborted fetal material through deoxyribonucleic acid (DNA) profiling. *Journal of Forensic Sciences* **36**:1219–1223.

Mange AP and Mange EJ (1980) *Genetics: Human Aspects* p. 495. Philadelphia: WB Saunders.

Marcotte A, Hoste B, Fays M, De Valck E and Leriche A (1996) Disaster victims identification by using the DNA technology and dental pulp: preliminary results. *Advances in Forensic Haemogenetics* **6**:334–336.

Martin P, Alonso A, Albarrán C and Sancho M (1996) Identification of the skeletal remains of two 12-years old bodies by nuclear DNA polymorphism analysis. *Advances in Forensic Haemogenetics* **6**:301–303.

Mevåg B, Jacobsen S and Olaisen B (1996) Three intriguing identification cases. *Advances in Forensic Haemogenetics* **6**:310–312.

Neuweiler J, Venturini J, Balazs I, Kornher S, Hintz D and Victor J (1992) The use of a chemiluminescent detection system for paternity and forensic testing. Verification of the reliability of oligonucleotide-probes used for genetic analysis. *Advances in Forensic Haemogenetics* **4**:132–136.

Odelberg SJ, Plaetke R, Eldridge JR *et al.* (1989) Characterization of eight VNTR loci by agarose gel electrophoresis. *Genomics* **5**:915–924.

Primorac D, Andelinovic S, Definis-Gojanovic M *et al.* (1996) Identification of war victims from mass graves in Croatia, Bosnia, and Herzegovina by the use of standard forensic methods and DNA typing. *Journal of Forensic Sciences* **41**:891–894.

Sakurada M, Honda K, Nakatome M, Tun Z, Shimada S and Misawa S (1998) Recovery of high quality DNA from 500-year-old skeletal remains for forensic DNA typing. *Progress in Forensic Genetics* **7**:150–152.

Sanz P, Prieto V and Andres MI (1996) PCR genotyping of dental pulp from old human skeletal remains and fire victims. *Advances in Forensic Haemogenetics* **6**:319–321.

Vogel F and Motulsky A (1979) *Human Genetics, Problems and Approaches*, p. 82. New York: Springer.

Walker RL (ed.) (1983) *Inclusion Probability in Parentage Testing*. Arlington, VA: American Association of Blood Banks.

Weber JL and Wong C (1993) Mutation of human short tandem repeats. *Human Molecular Genetics* **2**:1123–1128.

Wolff RK, Nakamura Y and White R (1988) Molecular characterization of a spontaneously generated new allele at a VNTR locus: no exchange of flanking DNA sequence. *Genomics* **3**:347–351.

Polymerase Chain Reaction

M S Schanfield, Analytical Genetic Testing Center, Denver, CO, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0469

Introduction

The introduction of DNA restriction fragment length polymorphism (RFLP) technology revolutionized the field of forensic identity testing. This is especially true for the area of sexual assault evidence that historically has been one of limited information. Although RFLP technology has been a tremendous aid, it has several problems. It is expensive to implement, labor intensive, expensive to test, and is limited by both quantity and quality of DNA obtained. The technical feasibility of amplifying specific segments of DNA using the polymerase chain reaction (PCR) has the potential to overcome the shortcomings of RFLP technology.

PCR-based technology is much less expensive to implement, as it does not require a laboratory capable of handling radioactive isotopes. It has higher productivity, in that it can be done more quickly, and each worker can do more cases in the same amount of time. PCR, by its nature, works with smaller amounts of DNA and with DNA that has been environmentally abused.

Polymerase Chain Reaction

At the same time as the revolution in RFLP was beginning in the mid-1980s, the development of modern PCR occurred. The role of the polymerase enzymes in copying DNA had been known since the 1970s. In early uses, the polymerase had to be replenished after the denaturation step because heating the polymerase to near boiling temperatures destroyed it. The use of polymerase obtained from bacteria living at high temperature (T*hermus* aq*uaticus* (*Taq*) polymerase) allowed for the automation of thermal cycling and the introduction of modern PCR. The Cetus Corporation initially developed modern PCR and obtained patents on the technology and the polymerase enzyme. They were interested in developing a better test procedure for doing crossmatching for transplantation. The region of interest in transplantation was the first PCR-based system to be applied to forensic testing. The marker was called DQα (now called DQA1), and consists of a sequence-based polymorphism.

Sequence-based polymorphisms are regions that code for proteins and are polymorphic. The copying of a simple polymorphism would mean that in one allele there is a C at a given position, and the other allele has an A. In a simple case, individuals would have two As, A and C or two Cs. Once the region has been amplified it has to be detected (see below).

Another type of polymorphism, first found using restriction enzymes, are length polymorphisms caused by variation in the number of repeated segments; these are referred to as restriction fragment length polymorphisms or RFLPs. The regions with variable number repeated segments are referred to as variable number tandem repeats or VNTR regions.

Definition and description of PCR

PCR is based on biochemical processes within cells that repair damaged DNA and make copies of the DNA as the cells replicate. In the repair mode, if a single strand of DNA is damaged, the damaged area is removed so that there is a single-stranded section of DNA with double-stranded sections at either end. The polymerase enzyme fills in the missing complementary DNA. In the copy mode an entire strand is copied during DNA replication. **Figure 1** illustrates a polymerase enzyme copying a portion of a strand of DNA.

In a cell a specific gene is copied or translated from DNA to RNA because the polymerase has specific start and stop signals coded into the DNA. To copy a sequence of DNA *in vitro*, artificial start and stop signals are needed. These signals can only be made once the sequence of the region to be amplified is known. Once a sequence is known, the area to be copied or amplified can be defined by a unique sequence of DNA. For a primer to recognize a unique sequence in the genome it must be long enough for no other sequence to match it by chance. This can usually be achieved with a sequence of 20–25 nucleotides. These manufactured pieces of DNA are called 'primers' and they are complementary to the start and stop areas defined above. The 'forward primer' is complementary to the beginning sequence on one strand of DNA, usually called the positive strand. The 'reverse primer' is complementary to the stop sequence on the opposite or negative strand of DNA.

Multiplexing PCR reactions

One of the advantages of PCR is that more than one region can be amplified at a time. Although it is necessary to select carefully primers that cannot bind to each other, the only limitation on how many pairs of primers are placed together is the ability to detect the amplified product. As we shall see below, multiplexing has advantages when testing for sequence or length polymorphisms.

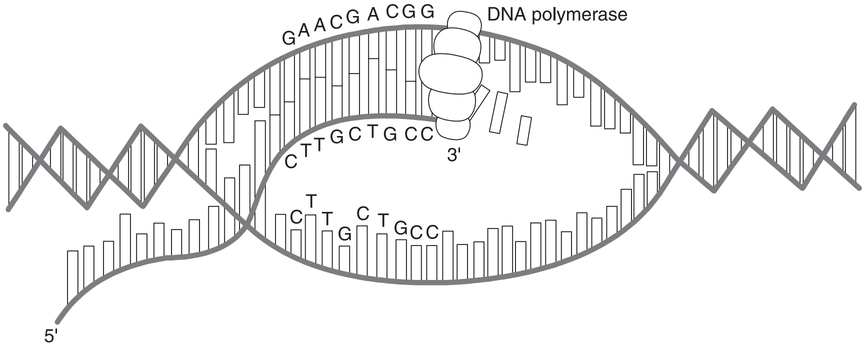


Figure 1 DNA polymerase copying one strand of a portion of double-stranded DNA.

PCR Process

To perform a PCR reaction several ingredients are needed. They include PCR reaction buffer, which is basically a salt solution at the right pH for the enzyme being used, the four nucleotides (DNA building blocks), primers, a thermostable DNA polymerase (Taq, Pfu, Vent, Replinase, etc.) and template DNA. The process consists of heating a solution of DNA to greater than 90°C. Double-stranded DNA comes apart or melts to form single-stranded DNA at this temperature. This is called the *denaturation* step. The solution is then cooled down to between 50 and 65°C so that the primers will bind to their complementary locations. This is called the *annealing* or probe hybridization step. Finally, the solution temperature is raised to 72°C, at which point the polymerase makes a copy of the target DNA defined by the primers. This is called the *extension* step. This completes one cycle of the PCR process. To make enough copies of the target DNA to detect, the process is repeated 25–40 times. This is done using a device called a thermal cycler. The process is illustrated in **Fig. 2**. If the process were perfect, 30 cycles would create over a billion copies of the original target DNA.

Thermal cyclers, in general, consist of an aluminum block with wells designed to fit the plastic PCR reaction tubes. The aluminum block has heating and cooling elements controlled by a microprocessor that can raise and lower the temperature of the block and the plastic PCR reaction tubes in the block. In the thermal cyclers that were first made, the plastic reaction tubes extended above the thermal block. This allowed cooling to take place above the reaction. The water in the reaction mixture would evaporate and condense at the top of the tube, changing the concentration of reactants and affecting the success of the amplification. To limit the evaporation, mineral oil was placed on top of the reaction mixture. New thermal cyclers have heated lids on top of the block to prevent or minimize evaporation. The microprocessor can store many sets of instructions, such that different programs can be kept to amplify different sequences of DNA.

Detection of PCR products

There are many methods for detecting PCR products. Since large amounts of product are produced, there is no need to use techniques such as radioactive detection, although it has been used in some clinical settings. In forensic testing, one of the advantages of PCR-based testing is that it does not require the use of hazardous materials. There is normally enough

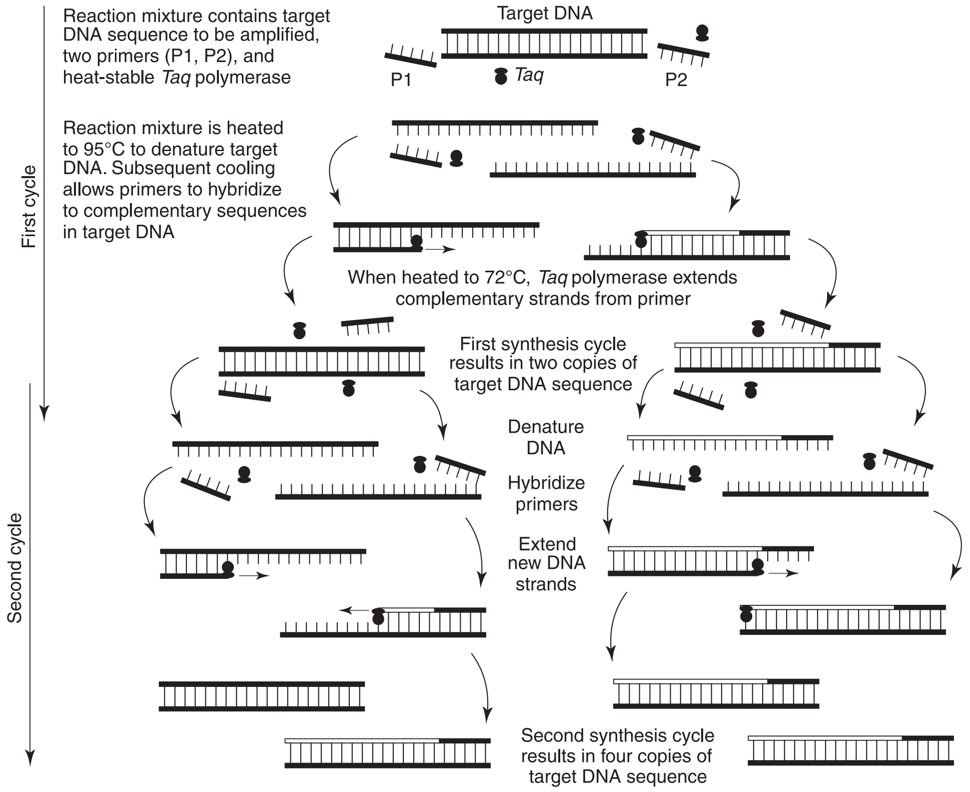


Figure 2 The PCR process. Courtesy of PE Cetus Instruments.

product so that if the PCR products are run on a yield gel and stained with ethidium bromide or Sybr green, there is normally enough DNA for detection. This is a suggested method to verify if the PCR amplification was successful and there is PCR product to detect.

Historical Perspective

Before discussing the specifics of detecting different PCR-based systems, a historical perspective is needed. The first typing system to be used forensically was the PCR test for HLA-DQα (now referred to as DQA1) region developed by the Cetus Corporation (and sold to Roche Molecular Systems) and marketed through the Perkin Elmer Corporation. This was one of the original PCR-based typing systems to be used for organ transplantation. In 1985, by historical chance Dr Edward Blake's laboratory was in the same building as the Cetus Corporation. He went to Dr Henry Ehrlich of the Cetus Corporation to see if PCR could assist him in solving a case involving a skull and a missing child. This was the first application of a PCR-based test to a forensic problem. Subsequently this was developed as the first commercially available PCR test to be used forensically. Although HLA DQA1 is informative, in approximately 16% of cases the victim and suspect (or multiple suspects) are of the same HLA-DQA1 phenotype. This can lead to difficulty in interpretation, especially if microscopic analysis of the extracts is not performed. Although DQA1 does provide information, the average level of individualization obtained using DQA1 alone is approximately 5% (**Table 1**). Therefore, there was a need for additional PCR-based systems to improve the discrimination of HLA-DQA1. PCR-based tests which evaluate variation in the sequence of the DNA are referred to as sequence-based polymorphisms (see below).

After its introduction, many different genetic typing systems were converted to PCR-based testing. Although RFLP-based tests were quite useful, some of the RFLP-based systems that had smaller fragments of DNA were investigated for the possibility of conversion to PCR-based testing. The first locus converted from an RFLP-based test system to a PCR-based test system was the 3′ hypervariable region of apolipoprotein B (APOB) by two laboratories in 1989 The term amplified fragment length polymorphism or AFLP, as the analog of RFLP, was coined. (Some individuals use the term 'AmpFLP'.) Since then, additional AFLP systems have been published and evaluated for forensic applications (D17S5/YNZ22, D1S80/pMCT118, COL2A1). The concept of an AFLP does not relate to the size of the VNTR: the early PCR-based systems had repeats ranging in size

Table 1 Observed likelihood of matching based on 19 European Americans.

	Polymarker							*Observed from data Likelihood of Match 1 in*			*Five RFLP loci*[a]
	DQA1	*D1S80*	*LDLR*	*GYPA*	*HBGG*	*D7S8*	*GC*	*DQA1*	*DQA1 + D1S80*	*DQA1+PM + D18S80*	
Median	0.0532	0.0341	0.3226	0.3114	0.4948	0.4792	0.317	19	632	328 074	15 887 773
Most common	0.0927	0.1792	0.4908	0.4933	0.4948	0.4792	0.3204	11	69	8 858	742 604
Least common	0.0076	0.0004	0.1866	0.1954	0.0028	0.1584	0.0784	132	73 099	73 841 357	1 243 444 270

DQA1, D1S80 and PM frequencies from FBI published databases.
[a] Five RFLP loci are D7S107, D11S129, D17S34, D18S17.2 and D21S112 with enzyme *Pst*I. They are less polymorphic than the *Hae*III systems in current use.

between 15 and 70. These are often referred to as large tandem repeats or LTRs. Once scientists started looking for length polymorphisms with PCR, many different polymorphisms were found – large numbers of dinucleotide (two repeats), as well as three-, four- and five-base pair repeat motifs. These smaller repeats are referred to as small tandem repeat or STR loci. All of these loci are AFLP based on the technology.

In the United States D1S80 was chosen for future development, and was ultimately produced as a kit. In Europe and elsewhere many different LTR AFLP systems have been used. D1S80, APOB and D17S5, also produced as kits, have been used the most extensively outside the United States. These systems are often still used in paternity testing.

Sequence Polymorphisms

The first PCR-based test used forensically amplified the DQA1 region of the human transplantation region. The system initially has six different alleles identified as DQA1*1.1, *1.2, *1.3, *2, *3 and *4. The *4 allele was subsequently divided into *4.1 and *4.2/4.3 in newer kits. These alleles differ in the sequence of the DNA and reflect differences originally detected at the protein level. To detect each of these different alleles a specific probe was produced that binds to a sequence representative of the different alleles. These probes are referred to as allele-specific oligonucleotide probes or ASOs. In general, each probe recognizes a sequence specific for different alleles. To create a commercial test the probes are bound to a strip of nylon membrane similar to that used in RFLP testing. This is called a reverse dot blot test.

Commercial kits were produced and distributed for DQA1 and D1S80. Research on the conversion of protein polymorphisms to PCR and the search for PCR-detectable sequence polymorphisms led to the production of a new kit, called Polymarker (PM), which consisted of five loci including LDLR, GYPA and D7S8 with two alleles, and HBGG and GC with three alleles. This was the first example of a multiplex PCR kit to be commercially manufactured. Eventually, the DQA1 kit and PM kits were merged to form a single kit which simultaneously amplifies six sequence-based loci, detected on two strips.

Reverse dot blot detection

To detect the PCR product for a reverse dot blot, the DNA is denatured by heating it to 95°C for 3–5 min to make the DNA single-stranded. The single-stranded PCR product is added to a probe strip in a detection tray. If a specific genetic type is present, that sequence will bind to the complementary sequence on the probe. This is called hybridization.

The hybridized strips are washed at high stringency to remove any DNA that is not specifically bound to the probe.

To allow for the detection of the PCR product bound to the probe, biotin is placed on one of the primers. Biotin or vitamin H has naturally occurring antagonists, or compounds that will bind to biotin to make it inactive. These antagonists are found in many organisms and are referred to by names such as anti-biotin, avidin or streptavidin (**Fig. 3**), depending on the organism it comes from. It is possible to put enzymes on to the streptavidin that will make certain dyes change colors under appropriate chemical conditions. These dyes will go from colorless and soluble to a colored, insoluble product that precipitates wherever the DNA has bound. Thus, after any nonspecifically bound DNA is washed off, the membranes are rinsed with a buffer. The prerinse in buffer for the enzyme substrate reaction helps to guarantee that the test will work. The dye in a buffer that will change color if the enzyme is present is added to the membrane strips. Wherever PCR product has bound to the membrane, there is enzyme bound to the biotin on the primer by the streptavidin molecules. Wherever there is enzyme, the dye will change color, creating a series of colored dots. An example of DQA1 strips is found in **Fig. 4**, and an example of PM strips is seen in **Fig. 5**. **Figure 6** is a flow chart of the entire procedure.

Each dot is labeled with an allele identifier to assist in the interpretation of the dots. In the case of DQA1 (Fig. 4) the left side of the strip has the nominal alleles 1, 2, 3 and 4; then there is a C dot and the subtyping dots which separate the 1.1, 1.2, 1.3, 4.1 and 4.2/4.3 alleles. The C dot is a probe that detects a nonpolymorphic segment of DQA1 and ensures that there was enough amplified DNA to yield an interpretable result. The S dot serves the same function on the PM strips. To interpret the DQA1 strips each ASO that changes color produces part of the type. For example, the victim (V) in Fig. 4 has only a 4 dot on the left, indicating that the donor is type 4. The subtyping dots on the right indicate that the sample is also 4.1- and 4.2/4.3-positive. The fact that the 1.2, 1.3, 4 dot and all but 1.3 dot are positive confirms the results. For suspect 1 (S_1) the 1 and 3 dots are colored on the left, indicating a 1,3 basic type. On the right the subtype 1.1 dot is colored, indicating the type to be 1.1,3. Suspect 2 (S_2) types as a 4.1 homozygote.

Interpreting the PM strips is easier. Each locus has two or three dots. Results are reported based on the dots that are colored. The top strip in Fig. 5 is type LDLR A, GYPA A,B, HBGG A,B, D7S8 A and GC A. In contrast, the bottom strip is LDLR A,B, GYPA B, HBGG A,B, D7S8 A,B and GC A,C.

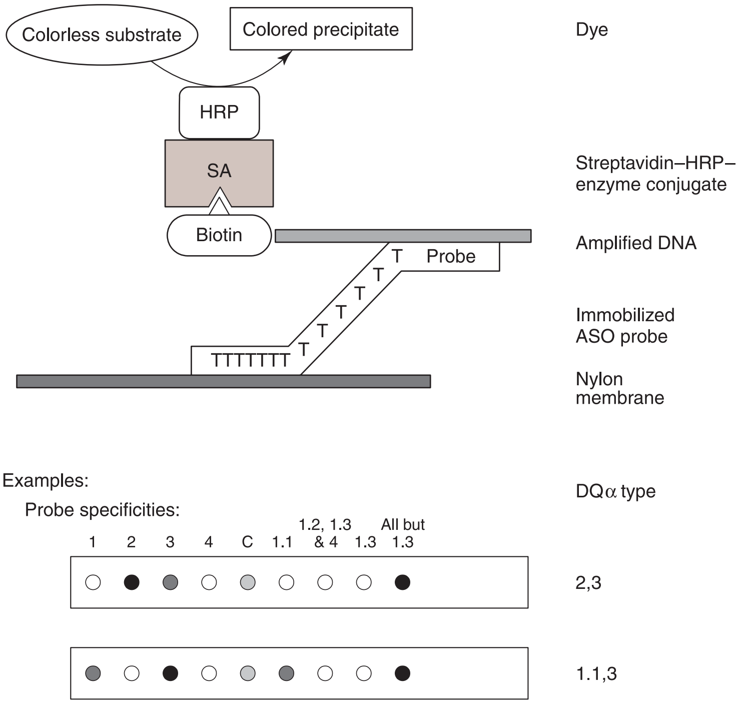


Figure 3 Generation of a reverse dot blot using immobilized allele-specific oligonucleotide (ASO) probes. Courtesy of PE Biosystems. HRP, horse radish peroxidase; SA, streptavidin.

Detecting AFLP loci

There are many different procedures for detecting length polymorphisms. At present, all of them require that the different sized fragments are separated by some form of electrophoresis. LTR AFLP loci can be separated on either agarose or acrylamide gels. Once the fragments have been separated by size they can be detected in basically two ways. Silver-staining stains the DNA with silver so that it turns a sepia color. This procedure is similar to the old methods of developing photographs. Unfortunately, silver-staining generates hazardous wastes, dangerous to personnel and expensive to dispose of. Further, the procedure is labor-intensive compared to fluorescence. **Figure 7** is an example of a silver-stained D1S80 gel.

The other method used to detect D1S80 after electrophoresis is direct fluorescence of the bands.

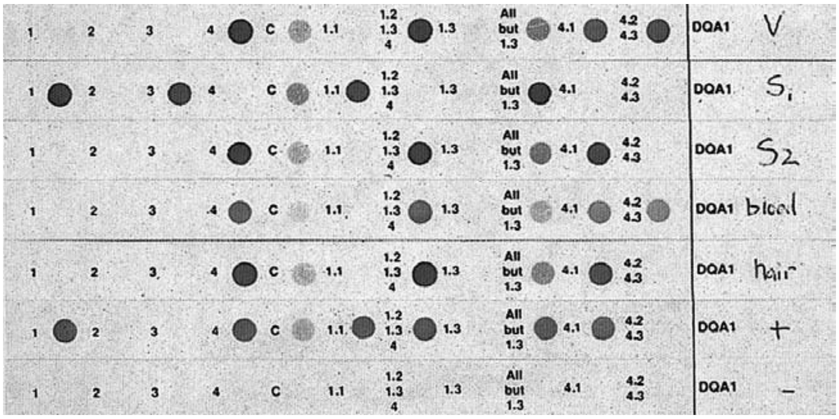


Figure 4 (see color plate 17) DQA1 strips representing a victim (V), two suspects (S_1, S_2), an evidentiary bloodstain, an evidentiary hair and positive (+) and negative (−) controls. The types for the victim and two suspects are 4.1,4.2/4.3, 1.1,3 and 4.1, respectively. The bloodstain is of the same type as the victim (4.1,4.2/4.3); the hair is of the same type as suspect 2 (4.1). Courtesy of PE Biosystems.

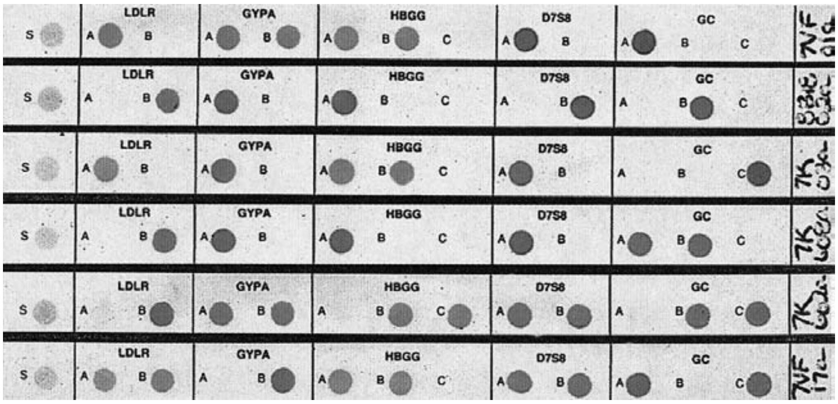


Figure 5 Examples of Polymarker (PM) strips demonstrating the different types at each locus. Courtesy of PE Biosystems.

The DNA in the gel is stained with a dye that binds to the DNA and fluoresces under exposure to certain types of light. Dyes such as ethidium bromide or Sybr green will bind to double-stranded DNA and fluoresce when exposed to ultraviolet light. These dyes can be applied directly to a gel. The simplest way of using fluorescent detection is to stain the gel and view it on a UV light box. A fluorescent scanner will also detect the fluorescent bands. A scanner has the advantage that images can be stored without photography and printed with a laser printer, enhanced so that weak bands are more readily observed.

Another procedure which is amenable to the multiplexing of AFLP is the use of fluorescent labels on the primers so that the amplified product will fluoresce when exposed to the appropriate light source. The amplified products can be electrophoresed and detected as described above. Fluorescent imaging also appears to be more sensitive than silver-staining and easier to perform. **Figure 8** is an example of fluorescent detection of a gel stained with Sybr green, detected on a fluorescent scanner. The band intensity is greater and the allele ladders appear darker. Both Fig. 7 and Fig. 8 tests were done with PE D1S80 kits.

Interpreting AFLP results

Sizing of RFLP fragments is done using a sizing ladder made up of fragments of viral DNA of known length,

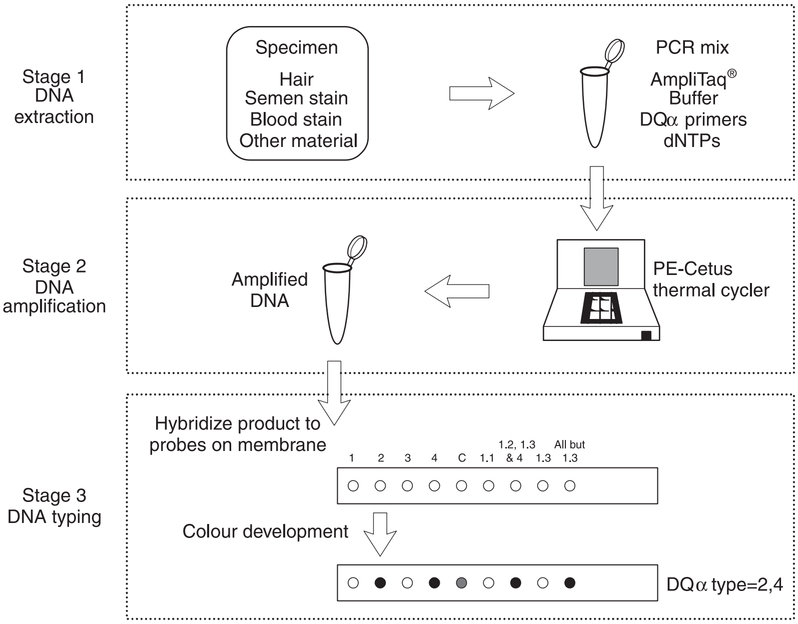


Figure 6 Flow chart for DQA1 typing of evidence. Courtesy of PE Biosystems.

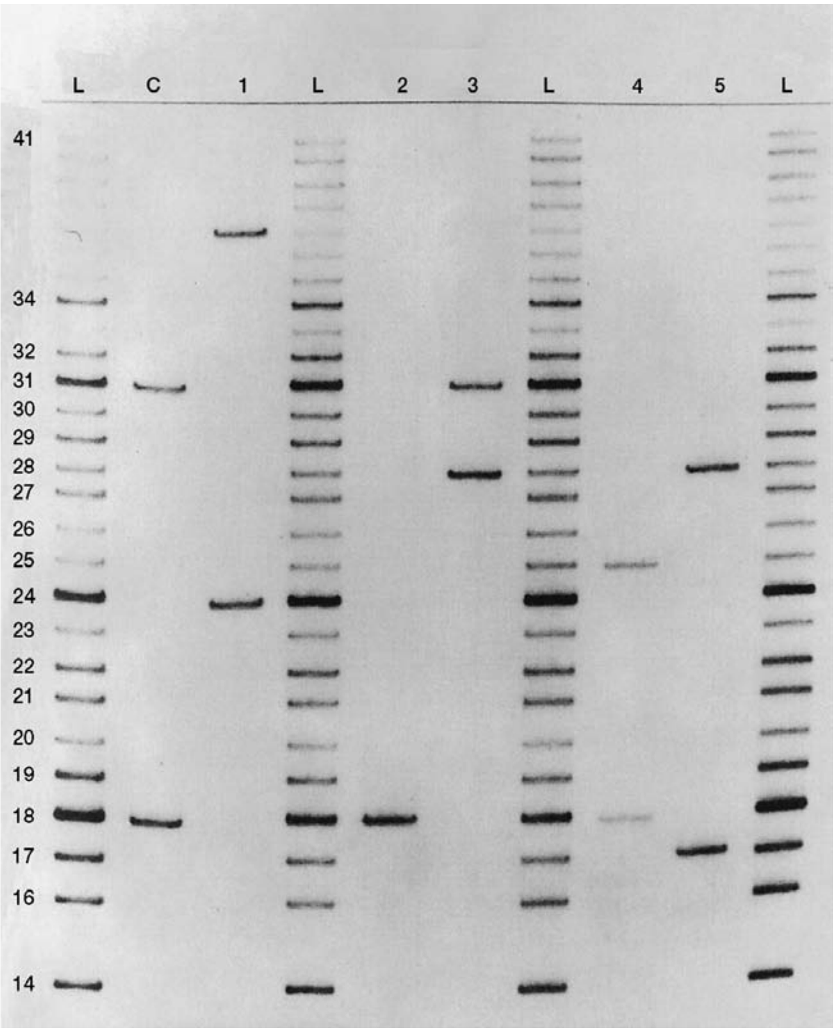


Figure 7 Silver-stained D1S80 gel. Types are as follows: lane C (control) 18,31; lane 1 24,37; lane 2 18; lane 3 28, 31; lane 4 18,25; lane 5 17,28. Lanes with L indicate allele ladders. Courtesy PE Biosystems.

created by cutting viral DNA, usually Lambda with various restriction enzymes and cocktailing the restricted fragments. A major source of variation in the reported sizes of RFLP DNA is due to variation in the sizing process and not to inherent differences in fragment size. Similarly, for AFLP fragments it has been observed that the ability to do both within and between gel comparisons of alleles is significantly hampered if sizing with conventional sizing ladders rather than allele ladders is used.

Since the unknown alleles and allele ladder fragments normally have the same relative mobility within a gel and comparisons are always made to known alleles, within and between gel comparisons are readily performed. The same alleles amplified and tested on different gels are always identified correctly;

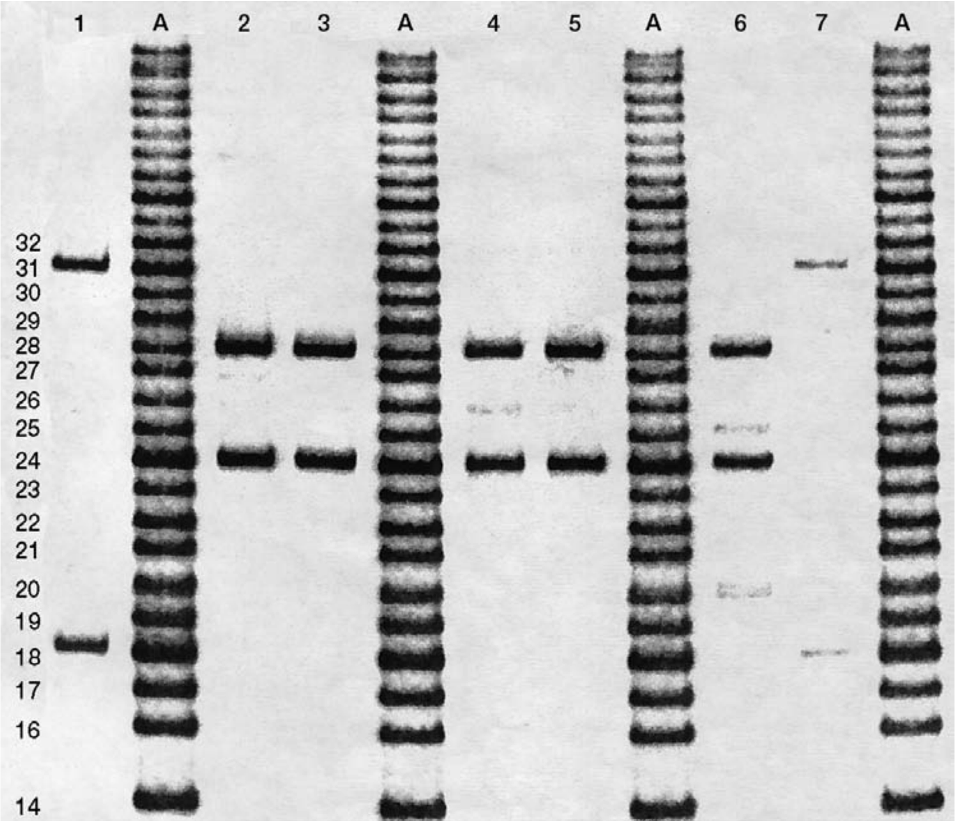


Figure 8 Fluorescently detected D1S80 gel. The gel was stained with Sybr green and imaged on a Genomyx SC after electrophoresis. Types are as follows: 1 18,31 (positive control); lanes 2, 3, 4, 5, 6 24,28; lane 7 18,31. Lanes 4, 5 and 6 contain mixtures. This is evidence from a case. Courtesy of Analytical Genetic Testing Center.

however, the sizings may not be the same. AFLP allele ladders are the combination of discrete component alleles with known VNTR repeat numbers, unique for each AFLP system. The system is similar to determining the types of the polymorphic enzyme PGM1, using a four-allele standard, usually referred to as a four-band standard. The only difference is a multifold increase in the number of alleles. Since migration is not totally linear in the range of the fragments observed, the different alleles automatically compensate for this variation. (Note that the bands with the higher numbers of repeats are slightly closer together than those with smaller numbers of repeats in (Fig. 7 and Fig. 8). For the D1S80 allele ladders in Fig. 7 and Fig. 8 the bottom allele is 14 repeats and the top rung is 41 repeats. In Fig. 7 the positive amplification control is immediately to the right of the far left allele ladder (lane identified as C), whereas in Fig. 8 the positive control is to the left of the far left allele ladder; in both cases the known type of 18,31 is readily determined. Published results of proficiency tests using allele ladders indicate that this is a highly reliable and reproducible method of typing samples, with no laboratory-to-laboratory variation in test results on the same samples.

Quality Assurance/Quality Control

Due to the sensitivity of PCR-based testing, it is necessary to go to extraordinary measures to make sure that accurate test results are obtained. It has been known for some time that the potential for contamination is much greater for PCR-based typing than any other type of testing. When PCR-based testing was first introduced there were articles on contamination of samples in the laboratory leading to unreliable results. Variation among laboratory personnel in the ability to work without contamination was such that, by 1993, guidelines were developed for the processing of samples for PCR-based testing for the general scientific community. In the forensic community the same types of guidelines were being developed. The Technical Working Group on DNA Analysis and Methods (TWGDAM) initially started working on developing guidelines for RFLP-based testing. They then continued and moved on to develop guidelines for PCR-based testing. Similar groups were meeting in Europe and Asia to develop standards for testing. The DNA Identification Act ultimately led to the DNA Advisory Board issuing guidelines for forensic DNA testing, effective 1 October 1998.

Even without these guidelines, which go beyond the scope of this article, the community of forensic scientists had set up procedures to minimize the possibility of contamination and maximize the likelihood of obtaining reliable results.

Sample collection

Samples collected at a crime scene are rarely pristine biological specimens. They are mixtures of environmental contaminants and biological fluids, such as blood, semen, vaginal secretions, saliva and urine, or any combination of these. This is expected and not a concern as far as contamination of specimens is concerned. Contamination of specimens refers to the introduction of foreign biological material after the evidence/samples have been collected. This could be the inadvertent mixing of body fluids because wet items of evidence were placed in the same container, or purified DNA introduced into an item of evidence. Unless the laboratory staff collect and package the crime scene material, they are not responsible for specimen contamination at the crime scene, only after it reaches the laboratory.

Although much of the following applies to the handling of all evidence, and may be covered elsewhere, some of these techniques are specific to PCR-based testing.

Laboratory structure

Different PCR functions should be isolated from each other. The basic separation of the evidence extraction area from the PCR setup area and the amplification area has been suggested and implemented in most laboratories.

DNA extraction area

General guidelines for the DNA extraction area include separation from other areas where DNA is handled. In some laboratories liquid blood samples (high concentration DNA) and dried evidentiary samples (low concentration DNA) are extracted at separate locations. DNA extraction should be carried out with dedicated laboratory equipment, such as pipettes, used in a PCR workstation that can be decontaminated with a built-in germicidal UV light (258 nm) capable of destroying DNA.

PCR extraction procedures

Regardless of the method used to extract DNA, whether organic, nonorganic Pro K, chelex, the same procedures are followed. Only a single item of evidence is open at a time. The tube receiving that sample for extraction is the only tube open at any one time. The tips used on the pipettes have a filter to block aerosol contamination. A tip is ejected after each use and a new tip applied to the pipette. For each set of samples extracted, the last extraction is an extraction control containing no sample, only the extraction chemicals. This is a control to ensure that the chemicals are not contaminated with DNA and thus adding DNA to the samples being tested. Another control added in some laboratories is called a positive extraction control or an extraction control. This is a biological material similar to that being extracted (liquid blood, dried blood, buccal swab), of known type, that is extracted and tested along with the evidence. It serves as a control to show that the extraction process is working.

Under ideal circumstances, the known samples (victim's and suspect's blood or buccal swabs) are not extracted at the same time as evidence. If that is not possible, evidence is always extracted before the known samples.

Following extraction, it is routine to quantitate the amount of DNA present. This is usually done using a slot blot that detects human DNA. PCR testing is sensitive to the amount of DNA. The amount of DNA to be added during the setup step is determined by the test being performed. In general, forensic tests have been designed to work well with 1–5 ng DNA. It should be noted that the most sensitive RFLP techniques usually will not go below 25 ng DNA. By knowing the concentration of DNA in the extracted sample, the volume of DNA to be added to the reaction mixture can be determined. If no DNA was detected, the maximum amount of sample allowed is added.

In some laboratories, dilute or low-level samples of DNA are concentrated using ultrafiltration devices or concentrators are used after quantitation to increase the chance of success.

PCR setup

Ideally the PCR setup area is in a different location from the extraction and amplification areas, or is an area that can be decontaminated before PCR setup occurs. Dedicated laboratory equipment is recommended and can be decontaminated after use. It is desirable to dispense the PCR reaction mixture into the reaction tubes first, as this minimizes the exposure of reagents to DNA and limits the number of pipette tips used. All of the tubes can then be closed and the DNA to be tested can be added to one tube at a time.

For the PCR setup two additional controls are added: (1) a positive amplification control, which is normally a sample of extracted DNA provided by the test manufacturer and of known type; and (2) a negative amplification control, which consists of the

PCR reaction chemicals and deionized water but no DNA. These are often set up last to show that none of the previous tests were contaminated by bad reagents.

Amplification and detection

The amplification and detection area should be the most isolated, as this area will contain amplified DNA in high concentration. Although no specific rules exist, it is sometimes more convenient to have the amplification area and detection area together. The amplified samples can thus be detected immediately and the product isolated or discarded when the detection is completed. It is generally considered desirable to keep the amplified product the most isolated, as it could be detected in a sample, even if it does not undergo additional amplification.

As with the other areas, this area should have dedicated pipettes and an area where the materials used can be cleaned and decontaminated. This area will contain thermal cyclers, shaking water baths for the detection of DQA1/PM strips and electrophoresis equipment to separate D1S80. If silver-staining is used to detect the D1S80, it will also require a certain amount of equipment.

Individualization of Evidence

Historically, DQA1 was the first PCR-based test system available to forensic scientists. Before the DQA1*4 allele was split into *4.1 and *4.2/4.3, approximately 16% of cases had two or more samples with the same DQA1 type. With the introduction of D1S80 the vast majority of these samples could be divided. The addition of PM made it possible to further individualize samples. To look at the discriminating power of these systems, profile frequencies were calculated for a series of samples that form part of a control panel. The genetic profiles of these samples are presented in Table 1. A series of 19 European Americans in a control panel have been extensively tested for the RFLP- and PCR-based tests. They can be considered to represent an estimate of the amount of information found in these loci for forensic purposes. The median value was chosen instead of the mean, as the median is less affected by extreme values. The calculated values represent the expected likelihood of a coincidental match if a sample had the median, most common or least common type at each locus. Contrast the distribution of results observed in the 19 individuals. Clearly the additions of D1S80 and PM increase the ability to discriminate samples. How do they compare to RFLP loci? The only data on these loci were for less polymorphic loci detected with the enzyme PstI. Therefore, the frequency of RFLP profiles should be considered to be very conservative. It is evident that five RFLP loci provide significantly more individualization power than the battery of DQA1, D1S80 and PM; however, there is still a great deal of information to be obtained from these systems until they can be replaced with more informative PCR-based systems.

See also: **Deoxyribonucleic Acid**: Restriction Fragment Length Polymorphism; Polymerase Chain Reaction-Short Tandem Repeats; Future Analytical Techniques. **Quality Assurance**: QA/QC.

Further Reading

Blake E, Mihalovich J, Higuchi R, Walsh PS and Ehrlich H (1992) Polymerase chain reaction (PCR) amplification of human leukocyte antigen (HLA)-DQ α oligonucleotide typing on biological evidence samples: casework experience. *Journal of Forensic Sciences* **37**:700–726.

Boerwinkle E, Xiong W, Fourest E and Chan L (1989) Rapid typing of tandemly repeated hypervariable loci by polymerase chain reaction: application to the apolipoprotein B3/hypervariable region. *Proceedings of the National Academy of Sciences of the USA* **86**:212–216.

Dieffenbach CW and Dveksler GS (1993) Setting up a PCR laboratory. *PCR Methods and Applications* **3**:2–7.

Horn GT, Richards B and Klinger KW (1989) Amplification of a highly polymorphic VNTR segment by polymerase chain reaction. *Nucleic Acids Research* **17**:2140.

Kasai K, Nakamura Y and White R (1990) Amplification of a variable number tandem repeats (VNTR) locus (pMCT118) by the polymerase chain reaction (PCR) and its application to forensic science. *Journal of Forensic Sciences* **35**:1196–1200.

Ludwig EH, Friedl W and McCarthy BJ (1989) High-resolution analysis of a hypervariable region in the human apolipoprotein B gene. *American Journal of Human Genetics* **45**:458–464.

Mullis K and Faloona F (1987) Specific synthesis of DNA. *in vitro* polymerase chain reaction. *Methods in Enzymology* **155**:335–350.

Priestley L, Kumar D and Sykes B (1990) Amplification of the COL2A1 3′ variable region used for segregation analysis in a family with Stickler syndrome. *Human Genetics* **85**:525–526.

Von Beroldingen CH, Blake ET, Higuchi R, Sensabaugh GF and Enrlich HA (1989) Applications of PCR to the analysis of biological evidence. In: Ehrlich HA (ed.) *PCR Technology, Principles and Applications of DNA*, pp. 209–223. New York: Stockton Press.

Wu S, Seino S and Bell GI (1990) Human collagen, type II, alpha 1, (COL2A1) gene: VNTR polymorphism detected by gene amplification. *Nucleic Acids Research* **18**:3102.

Polymerase Chain Reaction-Short Tandem Repeats

M S Schanfield, Analytical Genetic Testing Center, Denver, CO, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0470

Introduction

In 1989, at the same time that the first amplified fragment length polymorphism (AFLP) markers were first described using the polymerase chain reaction (PCR), a new class of repeats was also being reported. The large tandem repeat (LTR) polymorphisms had been referred to as minisattellite polymorphisms. The new class of di-, tri-, tetra- and pentanucleotide repeats were referred to as microsattelites. These markers became known as short tandem repeats (STRs). The amplification of LTR AFLP markers in general appeared to the relatively error free, with stutter bands or copying errors (extra bands in descending numbers of repeats) relatively infrequent. In actuality they do occur if a sensitive enough detection technique is used. In contrast, the newly discovered dinucleotide repeats were difficult to use because of the frequent occurrence of stutter bands. Three, four and five nucleotide repeat AFLPs have a decreasing frequency of stutter bands. The use of high base pair STR AFLP loci made it practical to use them for human gene mapping, parentage testing and identity purposes. As more of the markers became available, it was possible to choose those that were more polymorphic as well as those with different size fragments to begin making highly discriminating multiplexes. The use of fluorescent detection and automated equipment, originally for the sequencing of DNA, made it possible to create fluorescent multiplexes with automated detection. Unfortunately, it was some time before these discoveries could be moved into routine use. For purposes of simplicity the term 'STR' will be used to refer to any STR AFLP typing system.

Early History of Forensic STR Kits

Forensic and paternity testing laboratories became very interested in the possibility of using STRs for routine use. Many laboratories tried to develop their own multiplexes and detection systems. However, a patent was issued to Dr Caskey and the University of Texas, who in turn licensed all STR technology to the Promega Corporation, limiting the amount of outside work on the development of multiplexes. The widespread use of standardized kits had to wait for the Promega Corporation and PE Biosystems to release kits.

The first kits released by the Promega Corporation included single locus kits for the loci CSF1PO, F13A01, F13B, FESFPS, HPRTB, LPL, THO1, TPOX and vWA (incorrect nomenclature, should be VWA03, similarly to F13A01) and multiplex kits that could be detected using silver staining, including the CTT multiplex (CSF1PO, TPOX and THO1) and the FFV (F13A01, FESFPS, VWA03). Subsequently, a silver STR III system was released with the loci D16S539, D7S820 and D13S317. **Figure 1** (and see **Table 1**) is a silver-stained gel with CTT on the right side and FFV on the left side of the gel.

The next evolution was the introduction of fluorescently labeled primer multiplexes. The nature of fluorescence and fluorescent detection is described below. With the introduction and wider availability of devices capable of reading fluorescent gels, fluorescent STR technology became much more practical. The first device commercially available for visualizing multicolored fluorescent images was ABI Applied Biosystems model 373 sequencing device which used ABI's four-color sequencing technology (see below). (Note: Perkin Elmer Corporation subsequently acquired ABI and recently PE Applied Biosystems separated from Perkin Elmer. Products from this company will be referred to as PE Biosystems.) Subsequently, Molecular Dynamics, Hitachi, Pharmacia and others introduced devices that could visualize at least two colors. The ability to detect PCR STR multiplexes fluorescently simplified the detection from silver staining, and meant that multiplexes in at least two different colors could be detected, allowing for larger numbers of loci to combine. It also meant that machines could do the detection with minimal hands-on time compared with a detection procedure such as silver staining. From a forensic standpoint this means that, if the amount of DNA is limited, a single amplification could produce results on as many as nine or ten loci, generating population frequencies comparable with those obtained from restriction fragment length polymorphism (RFLP) testing with much less expenditure of time and staff. (See below for supporting information.)

The first generation of commercially available fluorescent multiplexes was produced by Promega and PE Biosystems. The first two fluorescent multiplexes released by Promega were labeled with fluoroscein (**Table 2**) and contained the loci CSF1PO, TPOX, THO1 and VWA03 (CTTV) in the first and F13A01, FESFPS, F13B and LPL (FFFL) (**Table 3**) in the second. PE Biosystems released two multiplexes, one labeled with the dye FAM called 'Green' and containing THO1, TPOX, CSF1PO and the sex

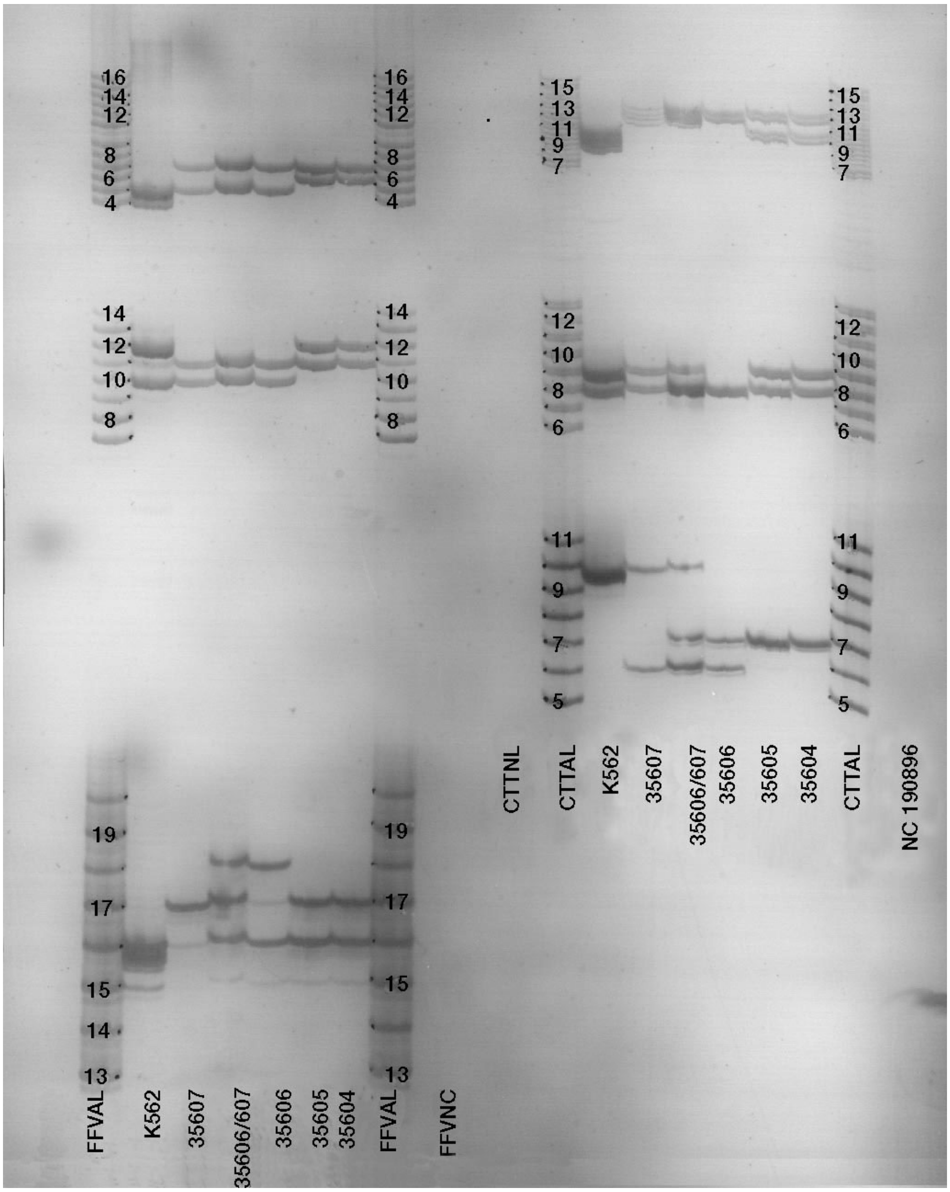


Figure 1 Silver-stained gel with both Promega CTT (right half of gel) and FFV (left half of the gel). CTTAL and FFVAL indicate the allele ladders. The lanes are: outside allele ladders; negative amplification control; positive amplification control (K562 cell line); 35607 (alleged father); mixture of 35607 and 35606; 35606 (child), and 35605 and 35604 (the mother). The results of the tests are presented in Table 1. The loci from top to bottom are CSF1PO, TPOX and THO1 on the right and F13A01, FESFPS, VWA03 on the left. Note: on the right in TPOX, where there is a mixture of 35606 and 35607, 35607 is type 8,9, 35606 is type 8, when the two are mixed the 8 band is much darker. The DNA is separated under conditions that make the DNA become single-stranded, forming doublets that can be seen at some of the loci. The VWA03 locus tends to form more pronounced stutter (repeat number minus one repeat bands) than the others; these are noticeable in the 'positive control' and other samples. At the VWA03 locus the child has an allele (*18) not found in the mother or the alleged father. This confirms an exclusion found previously.

identifier Amelogenin, and a second one labeled with JOE (**Table 4**) called 'Blue' and containing D3S1358, VWA03 and FGA. Subsequently, Promega released a third fluorescein-labeled quadplex called GammaSTR Multiplex that contained the loci D16S539, D7S820, D13S317 and D5S818 (Table 3).

PE Biosystems came out with a three-color kit called Profiler, which consisted of Green 1, Blue and a new multiplex called 'Yellow,' containing the dye TAMRA and the loci D5S818, D13S317 and D7S820 (Table 3).

In England the Forensic Science Service were also

Table 1 Results of silver stained CTT and FFV found in Figure 1.

Locus	*Negative Control*	*Positive Control*	*35607 AF*	*35607/6 Mixed*	*35606 Child*	*35605 Mother*	*35604 Mother*
CSF1PO	NR	9,10	11,12	11,12	12	10,12	10,12
TPOX	NR	8,9	8,9	8,(9)	8	8,9	8,9
THO1	NR	9.3	6,10	6,7,10	6,7	7	7
F13A01	NR	4,5	5,7	5,7	5,7	6,7	6,7
FESFPS	NR	10,12	10,11	10,11	10,11	11,12	11,12
VWA03	NR	16	17	16,17,18	16,18	16,17	16,17

AF, alleged father; NR, no reaction.
Mixed lane is a mixture of the alleged father and child's amplified product.

Table 2 Fluorescent dyes used to label primers for fluorescent STR detection

Common name	*Chemical name*	*Abs*	*Em*	*User*
Fluorescein	Fluorescein	494	518	Promega
5-FAM	5-Carboxyfluorescein	494 (493)	517 (522)	PE Biosystems (Blue)
JOE	Carboxyrhodamine	520 (528)	548 (554)	PE Biosystems (Green)
TAMRA	Carboxytetramethylrhodamine	542	568	PE Biosystems (Yellow, old)
NED	Unknown	(553)	(575)	PE Biosystems (Yellow, new
TMR	Tetramethylrhodamine	555	580	Promega
ROX	Carboxy-X-rhodamine	570 (587)	590 (607)	PE Biosystems (Red)
CRX	Carboxy-X-rhodamine	570	590	Promega

Values for Abs and Em taken from Haugland (1996); values in parentheses provided by PE Biosystems.
NED is proprietary. The technical service staff at PE Biosystems had no information other than Abs and Em. It appears to be in the TMR family of dyes.
Abs, absorption wave length; EMS, emission wave length.

Table 3 STR loci in Promega's fluorescent multiplex kits

STR locus	*Powerplex 1.1/1.2*	*Powerplex 2.1/2.2*	*Powerplex 16.2*	*CTTV*	*FFFL*	*Gamma*
D16S539	+ (fl)		+ (prop)			+ (fl)
D7S820	+ (fl)		+ (prop)			+ (fl)
D13S317	+ (fl)		+ (prop)			+ (fl)
D5S818	+ (fl)		+ (prop)			+ (fl)
CSF1PO	+ (TMR)		+ (prop)	+ (fl)		
TPOX	+ (TMR)	+ (TMR)	+ (TMR)	+ (fl)		
AMEL	+ (TMR)	+ (TMR)	+ (TMR)			
THO1	+ (TMR)	+ (fl)	+ (fl)	+ (fl)		
VWA	+ (TMR)	+ (TMR)	+ (TMR)	+ (fl)		
D18S51		+ (fl)	+ (fl)			
D21S11		+ (fl)	+ (fl)			
D3S1358		+ (fl)	+ (fl)			
FGA		+ (TMR)	+ (TMR)			
D8S1179		+ (TMR)	+ (TMR)			
Penta D			+ (prop)			
Penta E		+ (fl)	+ (fl)			
F13A01					+ (fl)	
FESFPS					+ (fl)	
F13B					+ (fl)	
LPL					+ (fl)	

Dyes used to label a multiplex are shown in parentheses: fl, fluorescein; TMR, tetramethylrhodamine; prop, a new dye from Promega.

Table 4 STR loci combined in each PE Applied Biosystems AmpF STR® kit

STR locus	*Dye color*	*Profiler*	*Profiler Plus*	*Cofiler*	*Blue*	*Green 1*
D3S1358	Blue	+	+	+	+	
VWA03	Blue	+	+		+	
FGA	Blue	+	+		+	
Amel	Green	+	+	+		+
THO1	Green	+		+		+
TPOX	Green	+		+		+
CSF1PO	Green	+		+		+
D5S818	Yellow	+	+			
D13S317	Yellow	+	+			
D7S820	Yellow	+	+	+		
D8S1179	Green		+			
D21S11	Green		+			
D18S51	Green		+			
D16S539	Blue			+		

Blue indicates the dye 5-FAM; Green indicates the dye JOE; Yellow indicates the dye TAMRA, later replaced with NED.

developing their own multiplexes to test evidence and profile criminals who had been convicted of significant crimes. The Royal Canadian Mounted Police (RCMP) in Canada were also looking at different multiplexes. Some of these multiplexes shared loci with those available in the United States; some were different.

At this point, with the large number of STR loci, a demand for a standardized panel in the United States and a need for there to be at least some sharing of loci with forensic counterparts in Canada, England and Europe, the Technical Working Group on DNA Analysis Methods (TWGDAM) implemented a multilaboratory evaluation of those STR loci available in kits in the United States. The loci chosen would be the PCR-based core of a national sex offender file required under the 1994 DNA Identification Act. The national program was initially called Combined DNA Indexing System or CODIS for short.

The TWGDAM/CODIS loci were announced at the Promega DNA Identification Symposium in the fall of 1997 and at the American Academy of Forensic Sciences meeting in February 1998. The following loci were chosen to be part of what was originally called the CODIS 13 loci: CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, FGA, THO1, TPOX and VWA03. These loci overlapped with the Forensic Science Services multiplexes and the Interpol multiplexes. The thirteen loci can be obtained in two amplifications using Profiler Plus and Cofiler (Table 3) from PE Biosystems, in two amplifications using Powerplex 1 and Powerplex 2 from Promega (Table 2), or in a single reaction with Powerplex 16, when it is released by Promega (Table 2).

Fluorescent Dyes

Before discussing the equipment used to detect fluorescent STRs, some understanding of fluorescent dyes is necessary. Fluorescent dyes or minerals when subjected to light at one wavelength, such as ultraviolet (UV) light or black light, will give off colored light at a slightly different wavelength or color. A characteristic of fluorescent dyes or materials is that the compound is excited at one frequency of light, referred to as its absorption or excitation peak, and emits or gives off light at a different frequency, referred to as its emission peak.

Ethidium bromide when bound to double-stranded DNA and exposed to a UV light source at 365 nm gives off pink light, whereas SYBER green gives off a green light. Ethidium bromide is about one-tenth as sensitive as SYBR green when using 365 nm UV light. This may be due to the fact that the peak absorption of ethidium bromide is 518 nm, with a peak emission at 605 nm (seen as pink to red light). The tail on the excitation side is quite broad so that there is some excitation at 365 nm. In contrast, the absorption peak for SYBR green is 494 nm, with an emission peak at 521 nm (seen as yellow to green light). This is much closer to the frequency of the light source, and therefore more able to use the energy coming out of the light source. Many different kinds of fluorescent probes that have uses other than labeling DNA have been described.

To label a PCR primer with a fluorescent dye, the dye is attached to the 5′ end of the molecule. Since DNA is translated from 5′ to 3′, it is at the very end of the primer and should not affect the amplification or binding of the primer if made correctly. One of the

oldest fluorescent dyes is fluorescein. Many devices have been made that will detect fluorescein, and it has been used extensively to label antibodies and other materials. Many of these dyes have been used for a long time and are in the public domain. Others have been developed for specific projects. The dyes used by PE Biosystems were originally proprietary and part of a patented four-color DNA sequencing system (Blue, Green, Yellow and Red). These dyes are now becoming more readily available. The dyes used commercially in the fluorescent detection of PCR-based STR systems are listed in Table 4, with the available information. The PE Biosystem Red (ROX) dye and CRX have not been mentioned. They are normally used for an internal size standard (see STR Detection, below).

Fluorescence Detection Equipment

The equipment used to detect the product of fluorescently labeled STR tests falls into two categories. The first are devices that scan a gel after it has been electrophoresed. Examples of this are the Hitachi FMBIO Fluorescent Scanner, the Molecular Dynamics Fluorimager and the Beckman Genomics SC scanner. The Hitachi and Molecular Dynamics scanners use a laser as a light, with filters to identify the proper frequency and a charge coupled device (CCD) camera to capture the image. The Beckman Genomics SC incorporates a monochromatic xenon light source and uses filters to detect the appropriate light for the CCD camera. The CCD camera scans back and forth over the gel as it is exposed to the light source and detects the various fluorescent colors using filters that change. This type of equipment has flexibility because different formats of electrophoresis gels can be used and scanned. The output is in the form of an electronic image with bands that look much like a set of RFLP bands. **Figure 2A** is an example of actual images (light on dark) recorded by the CCD camera, and Figure 2B is the reverse image (dark on light) that is reminiscent of an RFLP

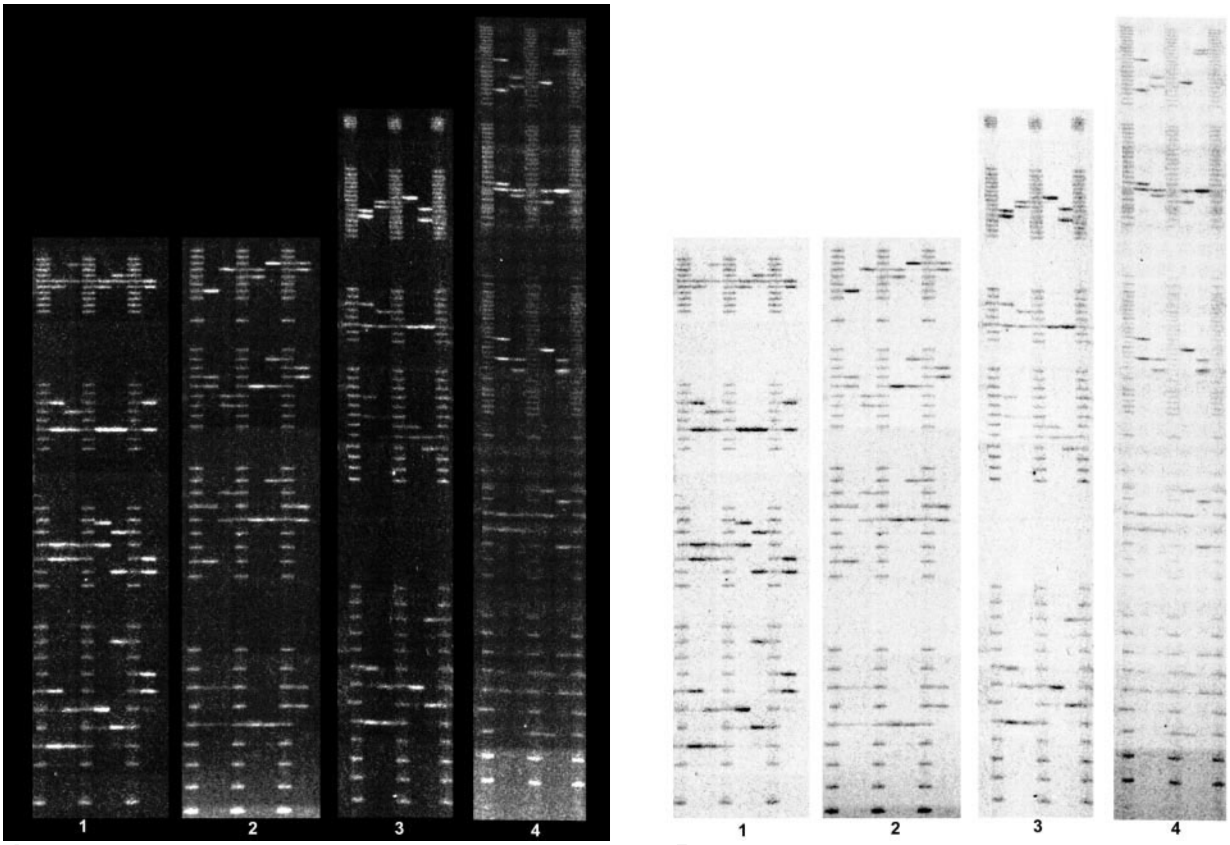


Figure 2 **(A)** The original CCD images of Powerplex 1.1 (panels 1 and 2) and Powerplex 2.1 (panels 3 and 4). **(B)** The reverse image of the same images. Panel 1 contains the TMR loci from top to bottom CSF1PO, TPOX, THO1 and VWA03. Panel 2 contains the fluoroscein loci from top to bottom D16S539, D7S820, D13S317 and D5S818. Panel 3 contains the TMR loci from top to bottom FGA, TPOX, D8S1179 and VWA03. Panel 4 contains the fluorescein loci from top to bottom: Penta E, D18S51, D21S11, THO1 and D3S1358. The same samples are loaded between the ladders. The sample format is the TWGDAM format when an internal size standard is not used. The overlapping loci for sample confirmation are THO1, TPOX and VWA03.

autoradiograph or lumigraph. It should be noted that the scanning time is added on to the electrophoresis time, with increased time for each color read.

The second type of imaging system is a real-time system, in which the DNA fragments, after the bands have been resolved, pass beneath a light source scanner that recovers the spectrum of light from the different fluorophors. This is the ABI Prism system from PE Biosystems. It includes the older model 373 DNA Sequencer and 377 DNA Sequencer, which use slab acrylamide electrophoresis to separate the DNA fragments, and the 310 Genetic Analyzer, which uses capillary electrophoresis to separate the DNA fragments. Capillary electrophoresis is a technology in which a fine glass capillary is filled with a proprietary separation polymer. The sample is pulled into the capillary by applying an electric current to it. Using high voltage electrophoresis (12 000 V), the DNA fragments are then separated over the length of the column and move past a laser detector. The 377 system can put approximately 60 samples on a gel at one time, and with modifications 96. In contrast, the 310 CE system does one sample at a time, with a separation time of approximately 20 min. However, as this is automated, a cassette can be filled with samples for testing and left to run unattended. The output of these devices is not a CCD image but a series of electropherograms with a profile for each color scanned (nominally Blue, Green, Yellow and Red). Since these are difficult to interpret, the computer software provides decomposed single color graphs. An example of Profiler Plus and Cofiler for a single sample is presented in **Fig. 3**. **Figure 4** contains an electropherogram of the single amplification Promega PowerPlex 16.2 System run on an ABI PRISM 310 Genetic Analyzer.

One of the extremely useful characteristics of fluorescent imaging devices is that the amount of light read by the detection device is quantified. The electropherograms produced by the real-time scanners is quantitative. However, the CCD image can also be used as a scanning densitometer to determine the amount of light in each peak. Since there is one fluorescent molecule per band, the amount of fluorescence is linear with the number of molecules in a band. This allows for many different types of analysis to be performed on the data generated. One of the more important of these is the ability to detect mixtures (see below).

PCR Setup and Detection

The manufacturers of the kits have done forensic validations of the kits, however each laboratory is responsible for the individualized validation required before testing. The guidelines for those validations for laboratories in the United States are governed by the DNA Advisory Board Guidelines (as of 1 October 1998), as implemented by ASCLD-LAB. Other areas of the world are regulated by other guidelines, unless they are also ASCLD-LAB accredited.

Virtually all of the procedures governing sample handling, contamination control and setup for the PCR apply to STR AFLP testing. Many of the procedures for STR typing are the same as those for DQA1, PM and D1S80 typing. The extraction procedure, quantitation, concentration (if needed) and setup are similar, if not identical. The reaction volumes for the two commercial STR kits vary. Promega's reaction is a 25 μl volume and limits the maximum sample to 10 μl, whereas PE Biosystems uses a 50 μl reaction volume and can take as much as 20 μl of sample. It may therefore be necessary to concentrate samples when using the Promega system. Amplification conditions are those stated in the manufacturers' protocols. The only major differences between DQA1, PM and D1S80 occur in the detection of the amplified product.

STR Detection

The major difference in the typing of the STR loci is the ability to include an internal size standard if the detection device used has multicolor capability. Under the TWGDAM guidelines forensic samples are to be placed adjacent to an allele ladder, as seen in Fig. 2. As the Beckman Genomics SC only has two filters (fluorescein and tetramethylrhodamine), an internal ladder could not be used, so the adjacent ladder format is used. In this situation there is no special preparation for detection. If the four-color Hitachi FMBIO II Fluorescent Scanner, or ABI Prism 377 or 310, and an internal standard are used the detection setup is modified slightly. As part of the set-up a ROX ladder for PE Biosystems kits (Fig. 3), or a CRX 60–500 bp ladder (Fig. 4) or a CRX 60–400 (**Fig. 5**) used with Promega kits, is added to all samples including the allelic size ladders. The internal size standard is used to size all fragments within a lane by detector supplied software and assigns repeat numbers from the allele ladder sizings.

Use of Internal Size Standards:

It was previously demonstrated that within gel variation in DNA migration could be compensated for by placing a size ladder within the lane and measuring each fragment with the internal size standard. This allows for highly precise measurements of fragments. Since the electrophoresis systems used to detect the STR loci have the capability of resolving as little as

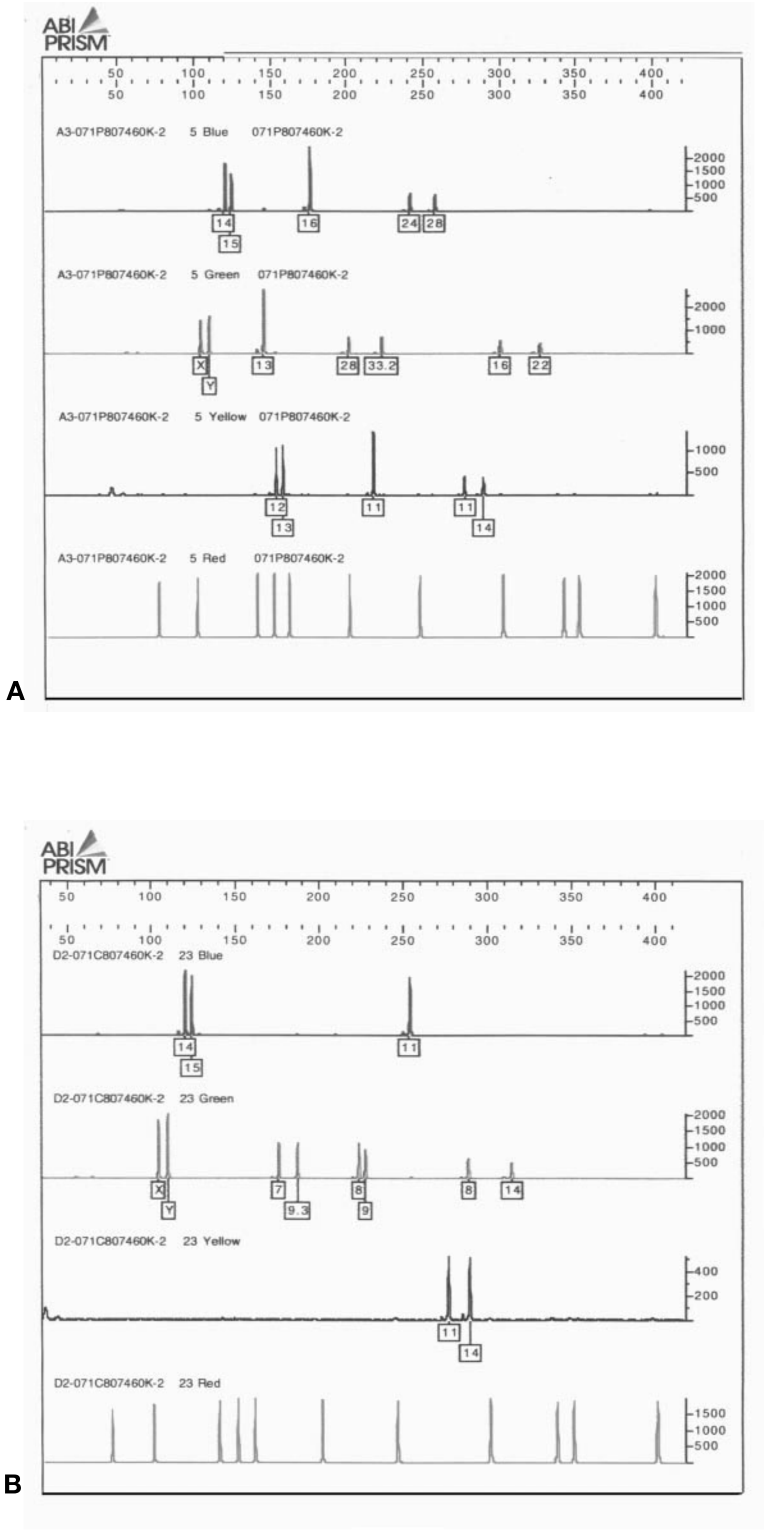


Figure 3 **(A)** Profiler Plus decomposed electropherogram. From left to right the loci are: Blue D3S1358, VWA03 and FGA; Green Amelogenin, D8S1179, D21S11 and D18S51; Yellow D5S818, D13S317 and D7S820. Red is the ROX sizing ladder. **(B)** Cofiler decomposed electropherogram. From left to right the loci are: Blue D3S1358 and D16S339; Green Amelogenin, THO1, TPOX and CSF1PO; Yellow D7S820. Red is the ROX sizing ladder. The overlapping loci for sample confirmation are Blue D3S1358, Green Amelogenin and Yellow D7S820.

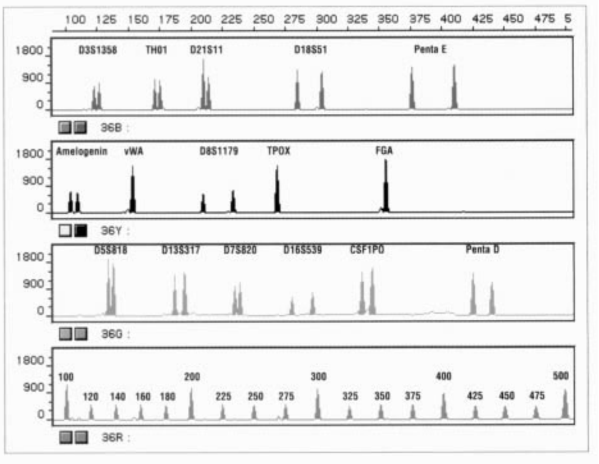


Figure 4 Electropherogram of a single DNA sample amplified using the 16-locus prototype Powerplex 16.2 System detected with the ABI PRISM 310 Genetic Analyzer. All 16 loci were amplified in a single reaction and detected in a single capillary. The fluorescein-labeled loci (D3S1358, THO1, D21S11, D18S51 and Penta E) are displayed in blue, the TMR labeled loci (Amelogenin, VWA03, D8S1179, TPOX and FGA) are displayed in black, and the loci labeled with a new dye (D5S818, D13S317, D7S820, D16S539, CSF1PO and Penta D) are displayed in green. The fragments of the prototype ILS-500 size marker are labeled with CRX and are shown in red. (Taken from Promega promotional literature.)

1 bp difference, it is necessary to have precise measurements to make sure that the fragment sizes can be accurately converted to repeat numbers. This would not be critical if all STRs were regular; that is, always have four base pairs for the tetranucleotide STRs. However, this is not the case. In THO1, one of the more common alleles, especially in Europe, is the *THO1**9.3* allele. Examples of this are seen in Fig. 1 (the positive control) and in Fig. 2B, panel 1, third sample from the left, in the third locus from the top. For other loci, such as FGA or D21S11, there are also intermediate repeats (see below).

Nature of STR Loci and Nomenclature

The loci DQA1, LDLR, GYPA, HBGG, D7S8 and D1S80 are all, with the exception of D1S80, sequence polymorphisms with defined alleles. The alleles have a simple nomenclature, or naming system. For example LDLR has two alleles, *LDLR*A* and *LDLR*B*; these can produce three phenotypes LDLR A, LDLR A,B and LDLR B. In the absence of family studies, we can only determine a phenotype. A phenotype is what is observed. A genotype is inferred from evidence that someone has two copies of an allele, or is homozygous. Thus in a family study it is possible to deduce that someone who is phenotype or type LDLR A is in fact *LDLR*A/*A*. When we are not doing family testing we cannot deduce this. Secondly, and perhaps more importantly, it is much easier to identify an apparent homozygote by using the type LDLR A.

The LTR AFLP D1S80 uses a nomenclature based on repeat units. It has been shown that people normally have between 14 and 41 repeats. Irregular repeats have not been found for D1S80. Occasionally, off-ladder electrophoretic variants alleles are encountered (alleles that do not line up with the allele ladder bands), but they usually reflect sequence changes within a repeat affecting electrophoretic properties on acrylamide gels and not variation in repeat size. In general, STR loci behave in the same fashion. However, irregular repeats do occur and the nomenclature has to compensate for that. A common allele at the THO1 locus has 10 repeats. Nine have 4 bp, but one has only 3 bp. Initially this allele was

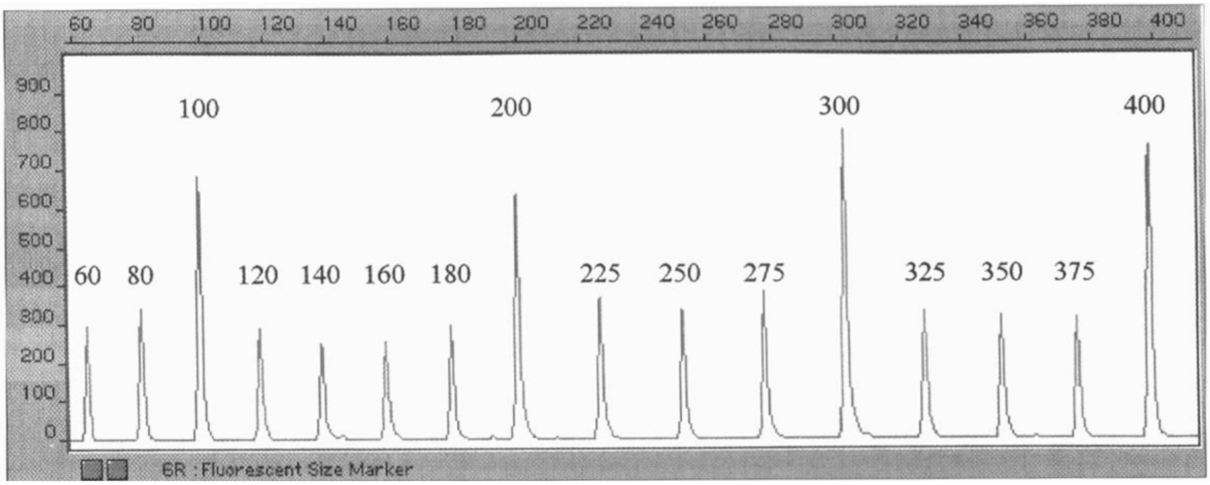


Figure 5 Electropherogram of Promega CRX labeled 60–400 bp ladder as visualized with the PE Biosystems ABI Prism 377 Sequencer. (Taken from Promega promotional literature.)

called 10-, which was not consistent with the International System of Genetic Nomenclature. The use of 9.3 indicates that there are nine regular repeats and one with 3 bp. The reporting of STR AFLP types are thus THO1 9, THO1 9,9.3 and THO1 9.3, to be analogous to the LDLR situation. These irregularities can take many different forms. Of the loci included in the CODIS 13, D18S15, D21S11, FGA and THO1, all have irregular repeat alleles. Most of them are .2 alleles, though .1 and .3 variants also occur – some with high frequency, e.g. *THO1**9.3*.

Usefulness in Detecting Mixtures

One of the major problems in the analysis of forensic evidence is posed by samples containing biological material from more than one source. Polymarker (PM) is very poor at sorting out mixtures because of the limited number of alleles present. DQA1 is slightly better, with its six alleles. D1S80 works well but is limited as a single LTR locus that does not often work well with degraded samples. In contrast, the larger number of discrete alleles at multiple loci make STR multiplexes an excellent tool for identifying components of mixtures. **Figure 6** shows the 'Green' loci from Profiler Plus from an item of case evidence indicating a mixture. With the quantitative aspects of the electropherogram it is possible to determine the major and minor contributors to the evidence. In the example in Fig. 6, the major contributor is female (X), D8S1179 10,15, D21S11 29,30 and D18S51 14,19, while the minor component is male (X,Y), D8S1179 13,?, D21S11 28,? and D18S51 12,15.

Statistics and Individualization

The calculation of population frequencies, likelihood of a match or probability of source are described elsewhere. The ability to calculate population frequencies for the individual phenotypes observed and the multilocus genotypic arrays all depend on the availability of databases. Often a laboratory will generate and use a database. In the United States the Federal Bureau of Investigation (FBI) recently published a validated database for the 13 CODIS loci. Most laboratories in the United States use this database.

To look at the level of individualization generated by the CODIS 13 loci, the same 19 individuals described in DNA (c) PCR for DQA1, PM, D1S80 and 5 RFLP loci were tested for the CODIS 13 loci (**Table 5**). The most common phenotype at each locus was detected for each of the loci. The actual 'least common phenotype', created by the co-occurrence of two rare alleles, was not detected for any of the loci, so the least common phenotype presented is that for the 19 samples and not the rarest combination one could find. The most common profile multilocus array represents the worst case scenario and the worst number obtainable if all loci were detected.

Individualization

In the United States the FBI has started releasing reports indicating that biological material originated from a specific source, much as fingerprint examiners have done for many years. The FBI has decided that if the population frequency exceeds 1 in 360 billion the sample is individualized. Other laboratories have chosen high threshold levels such as 1 in 600 billion. Whatever the level chosen, it is apparent that the median multilocus phenotype frequency exceeds both of these thresholds of individualization. The most common multilocus phenotype frequency does not exceed the threshold. None of the 19 individuals was under the individualization threshold of 3.6E+11 (360 billion). In fact the most common profile among the 19 Europeans was 3.113E+13, which is almost 100 times more than the threshold. Thus, although it is possible that individualization may not occur using the 13 CODIS loci, it appears to be relatively unlikely. The frequency of that most common phenotype occurring is the observed frequency, or about one in 200 billion individuals.

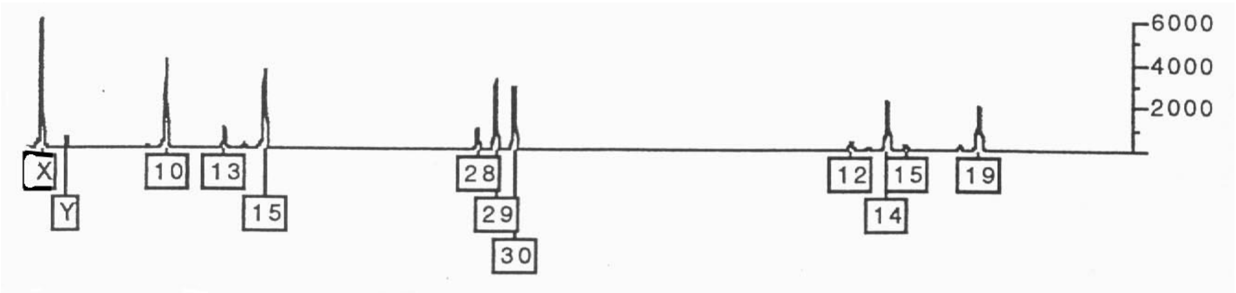


Figure 6 Green loci from a PE Biosystems ABI Prism 310 Genetic Analyzer on an item of evidence containing a mixture.

Table 5 Levels of individualization seen with three different test batteries in 19 European American individuals.

Test battery	*DQA1/PM/D1S80*	*5 RFLP loci*	*13 CODIS loci*
Median (*n* = 19)	328 000	15 900 000	5.028E+15
MCP	8 900	743 000	1.918E+11
LCP	73 800 000	1 240 000 000	3.788E+25

MCP, most common phenotype observed at each locus; LCP, least common phenotype observed at each locus.

See also: **Deoxyribonucleic Acid**: Basic Principles; Restriction Fragment Length Polymorphism; Polymerase Chain Reaction. **Evidence**: Statistical Interpretation of Evidence/Bayesian Analysis. **Identification/Individualization**: Overview and Meaning of ID.

Further Reading

Alford RL, Hammond HA, Coto I and Caskey CT (1994) Rapid and efficient resolution of parentage by amplification of short tandem repeats. *American Journal of Human Genetics* **55**:190–195.

Boerwinkle E, Xiong W, Fourest E and Chan L (1989) Rapid typing of tandemly repeated hypervariable loci by polymerase chain reaction: application to the apolipoprotein B 3′ hypervariable region. *Proceedings of the National Academy of Sciences of the USA* **86** 212–216.

Budowle B, Moretti TR, Baumstark AL, Defebaugh DA and Keys K (1999) Population data on the thirteen CODIS core short tandem repeat loci in African Americans, US Caucasians, Hispanics, Bahamians, Jamaicans and Trinidadians. *Journal of Forensic Sciences* **44**:1277–1286.

Edwards A, Civitello A, Hammond HA and Caskey CT (1991) DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *American Journal of Human Genetics* **49**:746–756.

Frégeau CJ and Fourney RM (1993) DNA typing with fluorescently tagged short tandem repeats: a sensitive and accurate approach to human identification. *BioTechniques* **15**:100–119.

Hammond HA, Jin L, Zhong Y, Caskey CT and Chakraborty R (1994) Evaluation of 13 short tandem repeat loci for use in personal identification applications. *American Journal of Human Genetics* **55**:175–189.

Haugland RP (1996) *Handbook of Fluorescent Probes and Research Chemicals*, 6th edn. Eugene, OR: Molecular Probes.

Kimpton KP, Gill P, Walton A, Urquhart A, Millinan EA and Adams M (1993) Automated DNA profiling employing multiplex amplification of short tandem repeat loci. *PCR Methods and Applications* **3**:13–22.

Litt M and Luty JA (1989) A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American Journal of Human Genetics* **44**:397–401.

Ludwig EH, Friedl W and McCarthy BJ (1989) High-resolution analysis of a hypervariable region in the human apolipoprotein B gene. *American Journal of Human Genetics* **45**:458–464.

Shows TB, McAlpine PJ, Bouchix C *et al.* (1987) Guidelines for human gene nomenclature. An international system for human gene nomenclature (ISGN, 1987). *Cytogenetics and Cell Genetics* **46**:11–28.

Weber JL and May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics* **44**:388–396.

Restriction Fragment Length Polymorphism

J S Waye, Department of Pathology and Molecular Medicine, McMaster University, Hamilton, ON, Canada

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0468

Introduction

More than twenty years have passed since the technique of Southern hybridization was first used to detect restriction fragment length polymorphism (RFLP) in human DNA, and a decade since this approach was used for human identification purposes in forensic casework. The RFLP method is a direct test in which genomic DNA is isolated from biological material and analyzed without any prior amplification via the polymerase chain reaction (PCR). Briefly, DNA is extracted and cleaved into small

fragments using a selected restriction endonuclease and polymorphic variation in the lengths of the resulting restriction fragments is used to distinguish between individuals. The analysis of a relatively small number (4–6) of highly polymorphic regions or loci within the human genome, commonly referred to as minisatellites or variable number of tandem repeats (VNTRs), is sufficient to generate DNA profiles that are virtually individual specific. Although the RFLP method is quickly being replaced by PCR-based typing methods, some forensic laboratories still rely heavily on RFLP typing. Moreover, RFLP typing continues to be a dominant technology for many paternity testing laboratories.

Genetic Markers Used for Forensic RFLP Typing

By far the most polymorphic sequences in the human genome are the so-called minisatellite or VNTR loci. As the name implies, VNTR loci are composed of short oligonucleotide repeat sequences which are organized into long tandem arrays. It is estimated that there are several thousand different VNTR loci in the human genome, representing a rich source of genetic variability. The length and sequence composition of the oligonucleotide repeats vary depending on the particular locus. For example, locus D1S7 is characterized by a 9 bp repeat, whereas D2S44 has a 31 bp repeat. The number of repeats in each array is highly variable and this polymorphism can be measured using a restriction endonuclease (RE) that has invariant or 'fixed' cleavage sites in the unique sequences immediately flanking the array. As shown in **Fig. 1**, the length of the resulting restriction fragment will be proportional to the number of repeats within the array.

The forensic application of RFLP typing came in 1985 when Alec Jeffreys and co-workers used the technique to generate individual-specific DNA fingerprints. This procedure uses hybridization conditions under which the VNTR probe is capable of simultaneously detecting many different VNTR loci. This approach, commonly referred to as multilocus analysis, results in complex and individual-specific hybridization patterns consisting of 8–33 bands per probe per individual. Although multilocus analysis affords a high degree of discriminating power, the complexity of the DNA fingerprints makes it difficult to delineate the possible origins of mixed samples, i.e. those containing DNA from more than one individual. In addition, multilocus analysis is not recommended for samples in which the genomic DNA is significantly degraded or limited in quantity.

In North America and elsewhere, forensic RFLP typing has been dominated by single-locus VNTR analysis. For this type of test, the VNTR probe is hybridized under conditions that preferentially restrict hybridization of the probe to its corresponding locus. The resulting hybridization pattern consists of one or two bands, depending on whether the individual is homozygous or heterozygous, respectively, for that particular locus. The successive typing of several independent VNTR loci enables one to assemble a multilocus DNA profile, one locus at a time (**Fig. 2**).

In an effort to develop standardized procedures and to facilitate the sharing of information, forensic laboratories have adopted protocols that are based on a single RE and a common panel of VNTR loci. For example, *Hae* III-based RFLP typing systems are used by the vast majority of forensic DNA typing laboratories in North America. *Hae* III recognizes the sequence 5′-GGCC-3′, cleaving between the second G

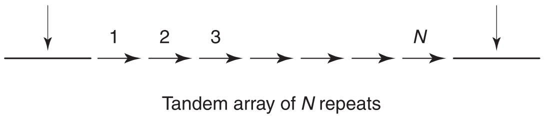


Figure 1 Schematic representation of a typical VNTR locus. The repeat sequence is such that *Hae* III does not cleave within it, but rather within the flanking unique sequences. Accordingly, the length of the *Hae* III fragment will be proportional to the number of repeats within the array.

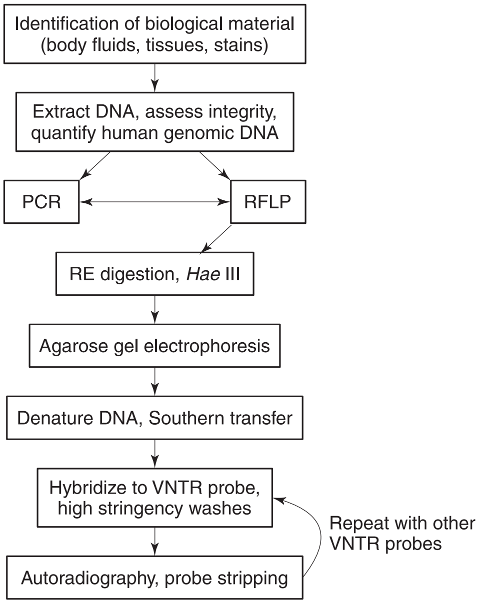


Figure 2 Schematic representation of the steps in the RFLP typing procedure using probes specific for multiple, independent VNTR loci.

and the first C and generating blunt-end fragments. *Hae* III cleaves the human genome into $>10^7$ fragments, most of which fall into the low-molecular-weight range of less than 10 kb. Several different VNTR loci are compatible for RFLP typing with *Hae* III, including D1S7, D2S44, D4S139, D5S110, D10S28, D14S13, D16S85 and D17S79. It is also noteworthy that forensic RFLP programs have been established using other REs (e.g., *Hinf* I and *Pst* I) and different VNTR loci.

In addition to VNTR loci, some forensic laboratories use probes specific for Y-chromosome sequences to determine the gender of the DNA source. One such locus, DYZ1, comprises several thousand copies of a tandemly repeated 3564 bp *Hae* III fragment. This fragment is detected in DNA from males (containing the X and Y sex chromosomes) but not in females (containing two copies of the X chromosome). As an internal control for the overall RFLP typing procedure, invariant or monomorphic loci can be analyzed. These are regions of DNA that yield the same pattern in all individuals, males and females alike. D7Z2 is a commonly used monomorphic locus composed of a tandemly repeated 2731 bp *Hae* III fragment.

The RFLP Typing Procedure

Extraction of genomic DNA

As with all DNA typing procedures, RFLP analysis begins with the extraction of genomic DNA from the biological samples of interest. Generalized extraction protocols can be applied to a wide range of body tissues and fluids, deposited on even a wider range of substrates. Initially, the cellular material in the sample is solubilized and lysed in a buffered solution containing salts, detergents, and proteinase K. The detergents lyse the cells and the proteinase K digests cellular proteins, including the histone and nonhistone components of chromosomal DNA. The resulting cellular debris is then removed by successive extractions with organic solvents (phenol, chloroform), after which the naked genomic DNA is purified by ethanol precipitation or filter microconcentration. As an alternative to proteinase K digestion and organic extractions, cellular proteins can be precipitated with high molarity salt solutions and the DNA purified by ethanol precipitation. Yet another approach involves the use of commercially available silica-based spin columns that selectively bind the DNA while excluding proteins and other cellular debris.

A subtle, yet very important modification of the above procedures is used to isolate genomic DNA from sexual assault swabs. Postcoital vaginal swabs generally contain both seminal fluid and epithelial cells, originating from the male and female, respectively. For such samples, the general DNA isolation procedure could yield a mixture of male and female genomic DNAs. This can be avoided by simple pretreatment of the swab to separate the intact sperm heads from other cellular material that may be present on the swab. The swab is first incubated in a solution containing detergents to lyse the vaginal epithelial cells and release the female DNA into the aqueous phase. However, since the lysis solution lacks reducing agents such as dithiothreitol (DTT), the sperm heads remain intact and can be pelleted by centrifugation. After differential separation, the sperm pellet can be processed in the presence of a reducing agent. By using such a strategy, the male (sperm) and female (vaginal epithelial) genomic DNA components of the sexual assault swab can be isolated and analyzed separately. It should be noted that the differential extraction of sperm and vaginal epithelial DNA from sexual assault swabs is not absolute. Depending on the relative proportions of sperm and vaginal epithelial cells or the degree of sample degradation, the male and female fractions may exhibit varying degrees of crosscontamination.

Assessment of DNA integrity and quantity

The overall quality or integrity of the genomic DNA is important since it is difficult to detect high-molecular-weight alleles from samples that are significantly degraded. DNA integrity can be easily assessed by agarose gel electrophoresis of a small portion of the unknown DNA sample and control samples of known DNA concentration and molecular weight. After electrophoresis, the yield gel is stained with ethidium bromide and the DNA is visualized by ultraviolet light fluorescence. The presence of a discrete high-molecular-weight band (>20 kb) on the yield gel is indicative of intact genomic DNA that may be suitable for RFLP typing. In contrast, degraded genomic DNA forms a continuum of fragment lengths (a smear), the appearance of which depends on the extent of degradation. Highly degraded samples appear as low-molecular-weight smears with little or no sequences above several kilobases in length, whereas moderately degraded samples appear as a high-molecular-weight fraction from which downward smearing is evident.

The yield gel provides a rapid and sensitive indication of DNA quantity and quality, but cannot distinguish human genomic DNA from that of nonhuman sources. Since forensic samples may be contaminated with nonhuman material (e.g. bacteria, fungi), additional tests can be used to quantify the amount of

human genomic DNA in a sample. This involves slot-blotting a small aliquot of the DNA sample to the surface of a nylon membrane, followed by hybridization with a labeled probe that recognizes highly repetitive human-specific DNA sequences (e.g. alpha satellite). The quantity of human genomic DNA in a sample is then estimated by comparing the intensity of its hybridization signal to that of known amounts of human genomic DNA. Once the amount of human genomic DNA has been accurately quantified, informed decisions can be made as to the amount of DNA required for successful RFLP typing. Although the RFLP method can be used to derive DNA profiles from as little as 50 ng of human genomic DNA, most forensic RFLP typing protocols have been optimized for amounts of genomic DNA on the order of 500 ng or greater.

Forensic RFLP typing

RFLP typing of VNTR loci involves a succession of analytical procedures (Fig. 2). The first step is to cleave the genomic DNA into small fragments with a restriction endonuclease (RE) such as *Hae* III. The lengths of the VNTR-containing fragments are then measured by sorting the DNA fragments according to size using a technique called agarose gel electrophoresis. DNA samples are loaded into wells along one end of the gel and an electric current is applied across the length of the gel. The DNA molecules migrate through the agarose gel, with the smaller fragments travelling at faster rates than larger fragments. After electrophoresis, the DNA fragments form an orderly array with the largest fragments located closest to the origin and a continuum of fragment lengths of decreasing sizes located at increasing distances from the origin. Each electrophoretic analysis also includes regularly spaced lanes of marker DNA fragments of known sizes. The sizing ladders are of nonhuman origin and therefore do not crosshybridize with human probes.

After electrophoretic separation, the double-stranded DNA fragments are denatured or converted into the single-stranded form by soaking the gel in an alkali solution. The fragile nature of the gel makes it unsuitable for the remaining steps in the RFLP procedure, a problem that is overcome by transferring the denatured DNA fragments from the gel to the surface of a nylon membrane (i.e. Southern transfer). The membrane-bound genomic DNA is then hybridized to a single-stranded DNA probe specific for the VNTR locus being examined. To detect the sizing ladder, labeled marker probe is also incorporated into the hybridization solution. The probes are labeled with a radioactive isotope of phosphorus (generally ^{32}P-dATP or ^{32}P-dCTP) or coupled to the enzyme alkaline phosphatase. After hybridization, the membrane is washed under high stringency conditions to remove any probe that may be bound nonspecifically to the membrane or to noncomplementary sequences.

Autoradiography is used to visualize the membrane-bound probe and thereby create a permanent image of the DNA profile. The membrane is overlayed with radiographic film and radioactive decay from the ^{32}P-labeled probe (or chemiluminescence from alkaline phosphatase labeled probes) gives rise to dark bands on the developed film, coinciding with the locations where the probe was bound to the membrane. The relative position of the bands is then used to estimate the length of the tandem array present at each VNTR locus. At each locus most individuals will be heterozygous, and the DNA profile will consist of two different fragment lengths appearing as two discrete bands located at different positions along the length of the membrane. A small minority of individuals will be homozygous, having single-band profiles because they inherited fragments of similar or identical size from both parents.

After the first VNTR locus has been analyzed, the probe is stripped from the membrane and the hybridization analysis is repeated using a probe specific for another VNTR locus. This process is repeated in a sequential fashion until a panel of four or more independent VNTR loci has been analyzed. **Fig. 3** shows RFLP profiles of four unrelated individuals derived using four independent VNTR loci. None of these individuals match at even a single locus, illustrating the high degree of discriminating power afforded by this type of analysis.

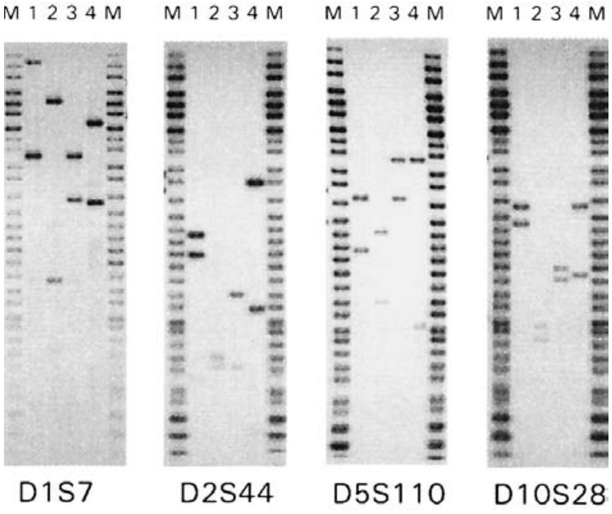


Figure 3 VNTR profiles of four unrelated individuals (lanes 1–4) derived using *Hae* III. The molecular weight marker (lanes M) is from Life Technologies.

Interpretation of RFLP Typing Results

Image analysis

Image analysis involves the examination of the hybridization pattern on the radiographic film to estimate the sizes of the corresponding DNA fragments (Fig. 3). The rate of electrophoretic migration for DNA fragments is inversely proportional to the logarithm of the molecular weight of the fragment. Applying this equation or derivatives thereof, a relationship can be drawn between the electrophoretic migration of the fragments of the sizing ladder and their known sizes. This in turn can be applied as a relative standard to estimate the size of any genomic DNA fragment from its electrophoretic mobility. Several different computer-assisted devices have been specifically designed for making objective determinations of the electrophoretic migration of sample fragments relative to a given sizing ladder.

Defining VNTR alleles

The degree of measurement imprecision associated with RFLP typing makes it impossible to treat VNTR fragments as truly discrete alleles. In recognition of this inherent technical limitation, conservative approaches have been developed for defining VNTR alleles and assigning allele frequencies. Many forensic laboratories use the so-called fixed-bin method to define VNTR alleles, in which the length of the electrophoretic separation is subdivided into 31 arbitrary size categories with defined boundaries. The dimensions of these size categories, or fixed-bins, are such that they adequately compensate for measurement imprecision. By using this method, continuous distributions of VNTR fragments can be treated as 31 arbitrarily defined alleles.

Another approach to estimate the frequency of a given fragment length involves the use of floating-bins, whereby conservative sizing windows are centered around the estimated size of the fragment of interest. For example, the frequency of a 5000 bp fragment would take into account the observed occurrence of fragments within a window of 5000 bp $\pm X\%$, where X is a value that reflects the empirically derived estimate of measurement imprecision for that particular laboratory. Floating-bins of $\pm 5\%$ are considered to provide frequency estimates that are as conservative as those derived using the fixed-bin approach. **Table 1** illustrates how the fixed-bin and floating-bin methods can be used to estimate the frequency of occurrence for a four-locus VNTR profile.

Prior to the most recent report of National Research Council, there was considerable scientific debate regarding the various methods used to estimate the frequencies of VNTR fragment lengths. However, both the fixed-bin and floating-bin ($\pm 5\%$) methods have been endorsed and continue to be used by forensic laboratories engaged in RFLP typing of VNTR loci.

Defining a match

As mentioned above, RFLP typing of highly polymorphic VNTR loci does not provide for the unambiguous resolution of individual fragment lengths. Due to the limited resolving capacity of the overall process, identical samples analyzed in adjacent sample lanes generally may not yield the same estimated fragment sizes (e.g. lane 1 = 2500 bp, lane 2 = 2400 bp). Thus, in an absolute sense, the estimated sizes of the VNTR fragments cannot be used to define a match.

Upon visual examination of two DNA profiles,

Table 1 Estimated frequencies of the four-locus DNA profile of individual no. 4 from Fig. 3. The frequencies were derived using a database for southern Ontario (Canada) and the fixed-bin and floating-bin (parentheses) methods. Floating-bins of $\pm 5\%$ were used for this analysis. All genotype frequencies were calculated using the formula $2pq$, where p and q are the frequencies of the two observed alleles. The estimated frequency of the four-locus profile is the product of the individual genotype frequencies

Locus	*Allele no. 1 (kb)*	*Frequency*	*Allele no. 2 (kb)*	*Frequency*	*Genotype frequency*
D1S7	6.73	0.0873 (0.0636)	3.65	0.0519 (0.0506)	9.062×10^{-3} (6.436×10^{-3})
D2S44	4.27	0.0192 (0.0183)	1.82	0.1154 (0.1154)	4.431×10^{-3} (4.224×10^{-3})
D5S110	5.34	0.0286 (0.0352)	1.73	0.0452 (0.0531)	2.585×10^{-3} (3.738×10^{-3})
D10S28	3.69	0.0666 (0.0440)	2.36	0.0745 (0.0316)	9.923×10^{-3} (2.781×10^{-3})
Estimated frequency of four-locus VNTR profile					1.030×10^{-9} (2.826×10^{-10})

there are three possible conclusions that can be made: (1) the profiles are so different as to warrant an outright exclusion; (2) the profiles match and an inclusion may be justified; or (3) no conclusions can be drawn from the existing data. In addition to visual assessment of the profiles, objective match criteria have been established for forensic comparisons. These criteria are based on the size estimates for the VNTR fragment lengths and involve an empirically derived match window within which two fragment lengths must fall before they can be declared a match. For example, the Federal Bureau of Investigation uses a window of $\pm 2.5\%$ to define the matches. Applying this match window, fragments of 2500 bp (2500 bp $\pm$ 2.5% = 2438–2562 bp) and 2400 bp (2400 bp $\pm$ 2.5% bp = 2340–2460 bp) could be declared a match because their size ranges are overlapping.

The significance of an RFLP match

Once a DNA match has been declared and the possibility of exclusion has been discounted, it is necessary to determine how common or rare the occurrence of this matching profile is in the population of individuals who could be the source of the evidence. Specifically, the relevant question becomes: what is the probability that someone other than the accused has a DNA profile that matches the evidence? To address this question, one can simply query a relevant population database and apply conventional genetic formulae to estimate the genotype frequencies for each matching locus. The product of these individual genotype frequencies can then be used to estimate the frequency of individuals who could fortuitously match at each of the loci tested.

The Hardy–Weinberg (H-W) equation is used to estimate the genotype frequencies for each individual locus. This simple binomial equation, $p^2 + q^2 + 2pq = 1$, can be used to estimate the frequency of homozygous (p^2, q^2) and heterozygous ($2pq$) genotypes based on the observed allele frequencies (p,q). For two-band VNTR profiles, it is assumed that the bands are allelic and that the individual is heterozygous. Applying the H-W equation, the genotype frequency for a two-band VNTR profile is calculated as $2pq$, where p and q are the estimated frequencies for the individual bands. For a single-band VNTR profile, it is reasonable to assume that the individual is homozygous and that p^2 can be used to estimate the genotype frequency. However, since it is formally possible that the individual may be heterozygous for a second undetected allele, many laboratories choose to express the genotype frequencies of single-band profiles as $2p$, where p is the estimated frequency of the detected allele.

Having estimated the genotype frequencies for the individual matching loci, the overall frequency for a multilocus DNA profile is calculated as the product of individual frequencies. The conservative nature of the binning procedure results in genotype frequencies for individual loci that generally are on the order of 0.01 (1/100). However, the application of the product rule across four or five loci results in overall frequencies that can be extremely discriminating (Table 1).

Strengths and weaknesses of RFLP typing

RFLP typing provides the capability to generate highly individual-specific DNA profiles through the direct analysis of only a few independent VNTR loci. Moreover, the relatively large amounts of DNA required for RFLP typing makes this method less vulnerable to contamination by DNA from extraneous sources (e.g. investigating officers or laboratory personnel). Notwithstanding the enormous impact RFLP typing has had on forensic science, there are several shortcomings associated with this methodology. First and foremost, the limited sensitivity of detection afforded by RFLP typing makes it difficult or impossible to obtain meaningful profiles from trace biological evidence or from samples that have significantly compromised due to aging or environmental insults. In addition, the interpretation of VNTR profiles can sometimes be difficult due to artifacts of the procedure such as partial *Hae* III digestion, band shifting, and incomplete profiles due to faint hybridization of low-molecular-weight alleles or degradation of high-molecular-weight alleles. Lastly, RFLP typing is quite time consuming, laborious and difficult to automate. Most, if not all of the inherent limitations of RFLP typing have been overcome through the development of PCR-based DNA typing systems.

See also: **Deoxyribonucleic Acid**: Future Analytical Techniques; Parentage Testing; Significance; Databanks.

Further Reading

Benzinger EA, Emerick E and Grigsby N *et al.* (1998) An illustrated guide to RFLP troubleshooting. *Journal of Forensic Sciences* **43**:665–679.

Budowle B and Baechtel FS (1990) Modifications to improve the effectiveness of restriction fragment length polymorphism typing. *Applied and Theoretical Electrophoresis* **1**:181–187.

Budowle B, Giusti AM and Waye JS *et al.* (1991) Fixed-bin analysis for statistical evaluation of continuous distributions of allelic data from VNTR loci, for use in forensic comparisons. *American Journal of Human Genetics* **48**:841–855.

Giusti AM and Budowle B (1995) A chemiluminescent-based detection system for human DNA quantitation and

restriction fragment length polymorphism (RFLP) analysis. *Applied and Theoretical Electrophoresis* 5:89–98.
Jeffreys AJ, Turner M and Debenham P (1991) The efficiency of multilocus DNA fingerprint probes for individualization and establishment of family relationships, determined from extensive casework. *American Journal of Human Genetics* 48:824–840.
National Research Council (1996) *The Evaluation of Forensic DNA Evidence. Committee on DNA Forensic Science: An Update*. Washington, DC: National Academy Press.
Waye JS and Fourney RM (1993) Forensic DNA typing of highly polymorphic VNTR loci. In: Saferstein R (ed.) *Forensic Science Handbook*, vol III. Englewood Cliffs, NJ: Prentice Hall.

Significance

A Carracedo and **M V Lareu**, University of Santiago de Compostela, Spain

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0473

Introduction

The discovery of polymorphisms in repetitive DNA has had a tremendous impact on Forensic Science. In the last ten years, highly informative and robust DNA typing systems have been developed which have proved to be very effective in the individualization of biological material of human origin.

DNA typing has advantages over traditional protein analysis. First it is more informative and can be analyzed in minute or degraded material, as DNA is physically much more resistant to degradation than proteins. Secondly, the same DNA genotype can be obtained from any tissue (i.e. blood, saliva, semen, hair, skin, bones), whereas the analysis of protein markers is restricted to cells where these proteins are expressed.

DNA analysis has become standard method in forensic genetics as it is currently applied by most labs for most of the forensic genetic types of expertise, especially in criminal forensic casework (stain analysis and hairs), identification and paternity testing.

PCR-based DNA typing systems have made it possible to analyze DNA obtained from only a few cells as well as from highly degraded human samples or remains, as has been recently demonstrated by the identification of relatively old human remains. An example is the identification of the remains of the Romanov family. The potential of DNA typing has also made possible the resolution of immigration problems or complicated paternity testing cases when the father is not available. Rapid identification of individuals in mass disaster using DNA typing has also been possible. Computerized DNA databases for the identification of criminal offenders have been created in some countries.

Due to all these impressive applications, the media have taken great interest in DNA profiling, the term firmly establishing itself in the everyday language of the man in the street, mainly because of the value of the evidence presented through DNA profiling in certain well-known legal cases.

Initially, the use of DNA profiling was very controversial in some countries, perhaps due to a hasty introduction of this new methodology. However this has ironically contributed to a much more reliable use of DNA profiling.

Two parallel upheavals concerning the introduction of DNA typing technology have accounted for this reliability: the introduction of quality control and accreditation schemes and in particular the spreading use of the statistics in the evaluation of DNA evidence. Although these two aspects began with conventional serological evidence, their importance nowadays is due to the great potential of this technology and the enormous value of the evidence that DNA profiling usually offers.

To understand the significance and value of DNA evidence, it is necessary to analyze the final consequences of the analysis. In some cases one DNA profile will not match another profile or a hypothesis, this event is called an exclusion, or that particular profile may match another or it may match a hypothesis (i.e. a parenthood relationship); in this case there is said to be a match. But how sure can these statements be? How infallible is an exclusion? And, is the DNA in the case of a match as good as a fingerprint? In addition, how can judges, lawyers, juries and the public in general be sure that a particular laboratory properly performs DNA analysis?

The aim in this article is to briefly address these questions.

The Value of an Exclusion

Criminal casework

When the DNA profile of a sample found at the scene of the crime does not match with the DNA profile of the suspect it can be concluded that both samples have a different biological origin. In most cases short tandem repeats (STRs) are used for DNA typing and

we can expect to obtain the same DNA profile from samples with the same biological origin even if different human tissues are used for comparison.

There are rare cases, particularly some human tumors (especially colorectal cancer) in which microsatellite instability can arise, so special care needs to be taken when using tissue cancer samples for identification purposes.

In general the '*a priori*' discrimination power of the STR used in casework is very high. With only a commercially available standard multiplex (**Fig. 1**) the discrimination power in Caucasians is higher than 0.99999. Most forensic genetic laboratories have a battery of DNA polymorphisms (including SLPs and STRs) of more than 15, so the theoretical discrimination power is enormous.

Generally, in polymerase chain reaction (PCR) based systems, reproducibility and standardization are made easier by using automated fragment separation and detection on automated sequencers and by applying a standardized system-specific sequenced allelic ladder as an internal measurement standard for allele designation. The use of sequenced allelic ladders is crucial for a robust typing.

The STR systems selected by any individual laboratory for casework should preferably be one of those currently used by forensic genetic laboratories and should have been previously validated. Laboratories and companies for the validation of individual STRs and kits have often followed SWGDAM (Scientific Working Group DNA Analysis Methods) validation schemes. In Europe during the last decade, EDNAP (European DNA Profiling Group) has performed an important work of validation of DNA polymorphisms from classical SLPs (single locus probes) analysis to STRs and mitochondrial DNA (mtDNA). For nomenclature the ISFH (International Society for Forensic Haemogenetics) recommendations are complied with by all the laboratories in the field.

Paternity testing

The mutation rate of DNA polymorphisms including both SLPs and STRs are being estimated. For STRs the mutation rates of different loci can differ by several orders of magnitude and even different alleles at one locus can exhibit different mutation rates. Factors affecting the mutation rate include the number of uninterrupted repeats, the size and the complexity of the STRs. Mutation events in the male germ line are five to six times more frequent for STRs than in the female germ line. The average mutation rate in the STRs can be estimated at around 1.2×10^{-3}.

In general, isolated exclusions by a single marker are obviously not considered as conclusive and the probability of paternity is then calculated excluding the marker having the possible mutational event. In general exclusions are considered in the report as conclusive when the exclusion number is more than three. In these cases the probability of a triple mutation has a chance on average below 10^{-8}. When we are more knowledgeable about the mutation rate of individual markers, it will then be possible to include in a report the theoretical error in case of exclusion.

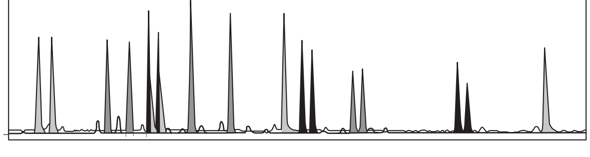


Figure 1 One of the most commonly used multiplexes (Profiler plus, PE) in forensic genetic laboratories. It contains nine STRs plus amelogenine. The discrimination power of this multiplex in Caucasians is 0.9999.

mtDNA

In the case of mtDNA when the patterns do not match by a single base pair (bp), the possibility of a mutation or heteroplasmy should be considered. The heteroplasmy and mutation rate depends on the methodology used. With very sensitive methodologies such as SSCP (single strand conformation polymorphism) or SSO (sequence specific oligonucleotides), the rate can be quite high. With sequencing, although the mutation rate is controversial, the possibility of a mutational event should be considered. In addition, there are sequence positions that have a higher probability of differing between known maternal relatives. There is also a handful of nucleotide positions that show a propensity for rapid substitution or the presence of heteroplasmy. One example of this is the position 16093, where more than 20 examples of substitution or heteroplasmy have been observed. Thus, if the only difference between two maternal relatives is at position 16093, the results should be at least considered inconclusive, and the data may be strong enough to give a verbal opinion in favor of a common biological origin if the accompanying polymorphisms are uncommon.

For all these reasons, very cautious verbal opinions should be given in the case of differences existing in one base, and the possibility that even in this case the samples can have the same origin should be considered. With differences in more than 3 bp (using sequencing) it seems highly likely that the samples have a different biological origin, but, again, cautious verbal opinions should be given and guidelines should be followed with the appropriate flexibility to account for the unique genetic characteristics of mtDNA.

The Value of a Match

Criminal casework

When two DNA profiles match, then the suspect is not excluded from being a contributor. In this case, it is necessary to attach some numerical weight to the evidence of a match, and statistical issues arise. The appropriate weighting for evidence is by means of a likelihood ratio. If *E* is the evidence of matching DNA profiles, and *C* ('the stain was left by the suspect') and *C* ('the stain was left by an unknown man') are alternative explanations for that evidence, then the relative merits of the two explanations can be compared by the ratio of the probabilities of the evidence under each hypothesis. It is then necessary to consider the probability of the evidence given in each of these hypotheses and the ratio of these two probabilities – the likelihood ratio – can be in the simplest cases to reduce to $1/f$, where f is the frequency of the observed genotype among members of the population to which the 'unknown man' credibly belongs. This must be estimated using data from a sample of people from the relevant population.

A STR profile is a multilocus genotype and probabilities of these profiles are estimated most simply as the products of the individual allele probabilities. The implied assumption of allelic independence can be tested with exact tests. At single loci, the independence assumption can be avoided by expressions that allow for the effects of population structure.

Nowadays, most experts use likelihood ratios for weighting the value of the evidence and for communicating this value to the courtroom. Frequencist approaches are prone to fallacies especially to the so-called prosecutor fallacy or the transposed conditional. Training of experts and judges on these concepts is necessary to prevent misconceptions when the evidence is presented in the courtroom.

There is a general agreement that the Bayesian approach to inference provides a logical and coherent framework for interpreting forensic transfer evidence. The interpretation of mixed stains is possible only in the context of likelihood ratios. Unlike single-contributor stains, the sample profile may not be certain under either of the two alternative propositions, so the likelihood ratio is the ratio of two probabilities that are less than one. Presenting the probability under only one proposition can be misleading.

Because of the enormous power of the evidence on many occasions provided by DNA analysis, the question often asked about a DNA profile is: Is it as good as a fingerprint? This apparently simple question has been addressed as follows:

> For non-DNA evidence (i.e. a fingerprint) individualization depends on a leap of faith that, in Bayesian terms is equivalent to saying: 'My personal likelihood ratio is so large that, no matter how small the prior odds are, the posterior odds are large enough to individualize with certainty'. For DNA evidence such a state could only be reached by testing more and more loci, but the apparent objectivity of numerical statements then becomes increasingly illusory, and the element of personal belief totally dominates the data. (Evett and Weir 1998).

However, if the likelihood ratio is enormous, expressions such as 'this profile is unique' or the mere introduction of the idea of uniqueness should be avoided and a numerical value given, otherwise the expert will be adopting the role of the judge.

Paternity testing

Parentage testing and identification of remains both exploit the genetic laws of the transmission of alleles from parent to child. As with forensic applications, DNA evidence is interpreted using likelihood ratios that compare the probabilities of the evidence under alternative propositions. In a normal paternity case the two propositions are as follows. H1: The alleged father is the father of the child; H2: some other man is the father of the child.

In general, high likelihood ratios are usually attained by DNA analysis in normal and even, but not always, in complicated cases. In complex paternity cases such as those associated with incest or paternity calculations when relatives but not the father are available, the correct consideration of the two alternative hypotheses is crucial.

Unlike the evaluation of the evidence in criminal casework, it is quite common for many experts to present the value of the evidence in terms of Bayesian probability using '*a priori* values' of 0.5. This is clearly wrong, since, at least in legal cases, the *a priori* value should be fixed by the judge and not by the expert.

Y chromosome polymorphisms and mtDNA

To evaluate the weight of the evidence in cases of matching using Y chromosome polymorphisms and mtDNA two problems arise. Firstly, paternally connected (using Y chromosome polymorphisms) or maternally connected individuals (in the case of mtDNA) will share identical haplotypes and therefore, cannot be discriminated. Secondly, the evaluation of evidence in terms of likelihood ratio presents some problems given that gene frequencies cannot be estimated using Hardy–Weinberg laws. In addition, population substructuring seems to be more severe in the case of Y chromosome markers than for unlinked autosomal markers. Due to this, large databases

including as many populations and individuals as possible are necessary for a more exact assessment. Taking these difficulties into account experts are, in general, giving verbal opinions in cases of matching with mtDNA and Y STRs instead of numerical values.

Communication of the value of the evidence in DNA analysis

Communication is closely linked to, and is as important as, interpretation. In general the most crucial factor in the communication of DNA evidence is not the explanation of the scientific techniques used and the explanation of the DNA results, but the communication of the value of the evidence. There are three main elements of an effective communication.

1. In all except the simplest cases, the interpretation of the scientific results depends on the circumstances, so it is necessary to detail these in the statement.
2. To assess the strength of the evidence it is necessary to consider at least two explanations for its occurrence. The evidence is evaluated by assessing its probability under each alternative hypothesis.
3. The strength of the evidence in relation to one of the explanations is the probability of the evidence given the explanation, divided by the probability of the evidence given the alternative hypothesis.

Some experts are in favor of giving verbal predicates in addition to the numerical values of likelihood ratios. Many laboratories use verbal predicates in paternity testing and verbal equivalents for LR. However, language is imprecise and the verbal predicates are merely a convention. More importantly, the expert is in some way taking the place of the judge when verbal predicates are used. The use of this approach may be understandable when juries are involved, for improving the comprehension of people who are unused to thinking numerically, but it seems an inappropriate approach for legal systems with professional judges and not juries.

Progress in Standards

If DNA analysis is nowadays accepted in all countries all over the world, it is in part due to the progress made in standardization.

Standardization of forensic DNA analysis has made enormous progress in the last few years and this innovation in standardization is comparable to the introduction of DNA technology itself.

Standards are crucial for forensic geneticists. This is due to the fact that only with an agreement about standards is it possible to develop quality control programs and quality assurance programs. In other words, standards are the only way to guarantee to judges, juries and the public that the tests performed and laboratory efficiency are reliable in any specific case. In addition, standards are necessary to allow for second opinions, to interchange data between labs and to create uniform searching procedures in cross border crime.

Two types of standards need to be addressed: technical and procedural. Technical standards include matters such as the genetic systems to be used (including type, nomenclature and methodology), the statistical methods for evaluating the evidence and the communication of the final report. Procedural standards encompass matters of operation, such as laboratory accreditation, laboratory performance, accreditation and licensing of personnel, record keeping and proficiency testing.

In the United States and in some European countries development of procedural standards for forensic genetics laboratories has made considerable progress in the last few years. In some of these countries laboratories have agreed on the requirements necessary for organization and management, personnel, facilities and security, evidence control, validation, analytical procedures, equipment calibration and maintenance, proficiency testing, corrective actions and audits. Proficiency testing programs for DNA analysis are established in some countries, and external and internal controls have been set up by most of the labs in western countries. Progress in accreditation has been effective in many countries in the last few years.

Even more advances have been made in attaining common technical standards. Agreement on genetic systems, types and nomenclature is widespread.

Establishing common standards in forensic DNA analysis is not easy due to the fact that there are very different legal systems and a variety of laboratories performing forensic genetic analysis. The success of the forensic geneticist in achieving common standards (at least compared with other aspects of forensic science and genetics) has been greatly facilitated by the ISFH, which has many national and international working groups, particularly EDNAP, actively involved in establishing common standards. In addition, the existence of commercially available kits for DNA typing, a shared aim of geneticists in the search for common standards and the influence of some leading groups, are other reasons for this success.

However, the efforts in standardization should continue. Progress in common procedural standards and particularly progress in similar requirements between countries for accreditation are necessary. With regard to technical standards other priorities

include the harmonization of criminal databases, the coordination and compilation of population databases (especially for mtDNA and Y STRs) and a continuation of progress initiated in the last few years on statistical evaluation and communication of the value of evidence provided by DNA analysis.

Conclusions

In the last ten years, highly informative and robust DNA typing systems have been developed which have proved to be very effective in the individualization of biological material of human origin.

DNA profiling has revolutionized forensic genetics and is widely accepted in legal cases. To make this possible, progress in standardization has been a crucial factor. Advances in establishing common procedural and technical standards have been considerable in the last ten years. Proficiency testing programs for DNA analysis have been established in some countries, and most laboratories have implemented external and internal controls. Progress made in accreditation has been effective in many countries in the last few years. Agreement on genetic systems, types and nomenclature is widespread.

In most cases, but not always, the value of the evidence provided by DNA analysis is enormous. However, uncertainty always exists. As scientists, we must measure this uncertainty and for this we use a standard: the probability. Likelihood ratios are nowadays used for weighting the value of the evidence and for communicating this value to the courtroom and the Bayesian approach to inference provides a coherent framework for interpretation.

See also: **Accreditation of Forensic Science Laboratories**. **Deoxyribonucleic Acid**: Restriction Fragment Length Polymorphism; Future Analytical Techniques; Parentage Testing; Statistical Analysis; Mitochondrial Deoxyribonucleic Acid. **Evidence**: Statistical Interpretation of Evidence/Bayesian Analysis. **Quality Assurance**: QA/QC.

Further Reading

Aitken CG and Stoney DA (1991) *The Use of Statistics in Forensic Science*. New York, London: Ellis Horwood.

Carracedo A (1996) *Problemas bioestadísticos en Genética forense*. Santiago de Compostela: USC.

Carracedo A, Brinkmann B and Bär A (1996) *Advances in Forensic Haemogenetics*, vol. 6. Berlin, Heidelberg: Springer.

Doutremepuich C (1998) *Les Empreintes Génétiques en Practique Judiciare*. Paris: La Documentation française.

Evett IW (1987) Bayesian inference and forensic science: problems and perspectives. *The Statistician* 3:99–105.

Evett IW and Weir BS (1998) *Interpreting DNA Evidence*. Sunderland, MA: Sinauer.

Holland MM and Parson W (1998) *Mitochondrial DNA Sequence Analysis in Forensic Casework: Methods and Current Issues*. Madison: Promega.

Lincoln PJ and Thompson J (1998) *Forensic DNA Profiling Protocols*. Totowa, NJ: Humana Press.

National Research Council (1996) *The Evaluation of Forensic DNA Evidence*. Washington: National Academy Press.

Olaisen B, Brinkmann B and Lincoln PJ (1998) *Progress in Forensic Genetics*. vol. 7. Amsterdam, Lausanne: Elsevier.

Statistical Analysis

B S Weir, North Carolina State University, Raleigh, NC, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0552

Introduction

Human individualization based on the genome exploits the fact that everyone, except for identical twins, is genetically distinguishable. Moreover, human genetic material is found in every nucleated cell in the body and can be recovered from samples as diverse as bone, bloodstains, earwax or shaver clippings. DNA may be recovered from very old samples that have been well preserved, and DNA signatures may even be preserved over successive generations.

The very genetic nature of DNA evidence that makes it of such value to forensic science also brings particular problems in interpretation. One problem is that as more and more genetic loci are used for DNA profiles, the proportion of a population who might be expected to share a particular profile becomes less and less. There comes a point where any attempt to quantify the probability of seeing a profile strains credulity. Certainly numbers such as 1 in 8 trillion fall into that range. Another problem is that the genetic constitution of a population is dependent on previous generations. Immediate family members are likely to have similar DNA profiles, but even apparently unrelated members of a population have a shared evolutionary history. Conveying the evidentiary strength of matching DNA profiles therefore requires the use of both probability and genetics.

Probability

Although the term 'probability' features prominently in describing DNA evidence, the meaning of the term is often left unclear. Indeed there are several possible approaches to defining probability. A classical definition is based on the notion of the proportion of times an event will occur in repeated opportunities for occurrence. When the probability that a well-shuffled deck of cards will have an ace on top is said to be 1 in 13, there is an implicit understanding that the event will be seen in one-thirteenth of many shufflings of the deck. When a weather forecaster says there is a 70% probability of rain that day, the meaning is less clear. Maybe the current conditions have led to rain in 70% of previous occasions, but the listener will need to make decisions about taking an umbrella based on his or her understanding of this figure. The personal element is even more clear in a courtroom when a prosecutor asks a jury to convict the defendant because the probability of guilt is so high as to be beyond reasonable doubt. The probabilities attached to DNA profiles cannot be interpreted as population proportions when they are very much smaller than one over the population size. They might be interpreted as personal measures of uncertainty, or they might be explained as being the result of a set of prescribed calculations. Even such calculations, however, necessarily have some subjectivity on the part of the person performing the calculations.

Laws of probability

However probability is defined, it must obey a set of rules in order to be useful. Suppose there is interest in some event H: that the card on top of a well-shuffled deck is an ace, or that it will rain today, or that the defendant is guilty. There is some information I about the event: the deck has four aces among the 52 cards, or the weather forecaster is very good, or the evidence against the defendant is very strong. The probability of H given the information I is written as $\Pr(H|I)$, and it satisfies three 'laws of probability'.

The first law says that probabilities are numbers between zero and one (inclusive), and that the probability of a certain event is one. In symbols:

$$0 \leq \Pr(H|I) \leq 1$$
$$\Pr(H|H) = 1$$

The second law says that the probability of either of two mutually exclusive events G, H happening is the sum of the separate probabilities for G and H. In the card example, the events G: 'the top card is a king' and H: 'the top card is an ace' are mutually exclusive because they cannot both occur. Likewise 'rain' and 'no rain' or 'guilty' and 'not guilty' are pairs of mutually exclusive events. In symbols:

$$\Pr(G \text{ or } H|I) = \Pr(G|I) + \Pr(H|I)$$

From the first law, this implies that the probability of defendant being either guilty or not guilty, given the evidence, is one.

The third law gives the probability of both of two events, G and H, occurring. The joint probability is the probability that first of them occurs multiplied by the probability of the second, given that the first has occurred. The 'first' and 'second' labels are arbitrary, so symbolically:

$$\begin{aligned}\Pr(G \text{ and } H|I) &= \Pr(G|I)\Pr(H|G,I)\\ &= \Pr(H|I)\Pr(G|H,I)\end{aligned}$$

For the card example, suppose G is 'the top card is a face card' and H is 'the top card is an ace'. For a standard deck of cards:

$$\Pr(G|I) = 16/52 \quad \Pr(H|G,I) = 1/4$$
$$\Pr(G \text{ and } H|I) = 1/13$$
$$\Pr(H|I) = 1/13 \quad \Pr(G|H,I) = 1$$
$$\Pr(H \text{ and } G|I) = 1/13$$

If knowledge that one event has occurred does not affect the probability of another event, the two events are said to be independent. When events G, H are independent:

$$\Pr(G \text{ and } H|I) = \Pr(G|I)\Pr(H|I)$$

Bayes' theorem

The laws of probability can be used to derive a result that is central to the interpretation of DNA evidence. From the third law, for any two events G, H:

$$\begin{aligned}\Pr(H|G,I) &= \frac{\Pr(H \text{ and } G|I)}{\Pr(G|I)}\\ &= \frac{\Pr(G|H,I)\Pr(H|I)}{\Pr(G|I)}\end{aligned}$$

The second expression is the simplest form of 'Bayes' theorem' and it allows the probability of an event given a second event to be expressed in terms of the probability of the second event given the first. In other words there has been a transposing of the conditional from $H|G$ to $G|H$. This transposition requires knowledge of the separate probabilities of G and H.

The law can be written in an 'odds form' by rewriting it for the event $\bar{H}$: 'not H':

$$\mathrm{P}(\bar{H}|G,I) = \frac{\Pr(G|\bar{H},I)\Pr(\bar{H}|I)}{\Pr(G|I)}$$

and then dividing the first expression by the second:

$$\frac{\Pr(H|G,I)}{\Pr(\bar{H}|G,I)} = \frac{\Pr(G|H,I)}{\Pr(G|\bar{H},I)} \times \frac{\Pr(H|I)}{\Pr(\bar{H}|I)}$$

The probability of an event divided by the probability of not that event is called the odds, so $\Pr(H|I)/\Pr(\bar{H}|I)$ is the 'prior odds' of H, i.e. the odds prior to knowledge of event G. After knowledge that event G has occurred, the ratio $\Pr(H|G,I)/\Pr(\bar{H}|G,I$ is the 'posterior odds' of H. In equation form:

$$\text{Posterior odds} = \text{LR} \times \text{Prior odds}$$

where the ratio of the two conditional probabilities $\Pr(G|H,I)$ to $\Pr(\bar{G}|H, I)$ has been termed the 'likelihood ratio' LR.

Forensic Probabilities

This formal language of probability theory has an immediate application for the interpretation of DNA evidence. Although the ultimate question in a trial concerns the guilt or innocence of the defendant, attention here is focused on just the DNA evidence. A common situation is where DNA is recovered from a biological sample left at the scene of a crime, and there is reason to believe the sample is from the perpetrator of the crime. DNA is also extracted from a blood or saliva sample from a suspect in the crime and is found to have the same profile as the crime sample. The events G and H are:

G: The DNA profiles of the suspect and crime sample (are the same).
H: The crime sample is from the suspect.
$\bar{H}$: The crime sample is from some other person.
I: Other information relevant to the probabilities of G and H.

If the suspect is charged with being the perpetrator, and becomes the defendant in a trial, the court is going to be interested in the probability of the defendant being the perpetrator given the evidence of a DNA match. Bayes' theorem makes it clear that this can be calculated only if some prior probability can be assigned. An extreme view might be that, prior to the DNA evidence, the defendant had the same probability as anyone else in the population of being the perpetrator. For a population of size N, the prior odds would be $[(1/N)]/[1 - (1/N)[= 1/(N-1) \approx 1/N$. The posterior odds would be LR/N.

It is not customary to present prior odds in criminal trials, although they are routinely used in civil paternity disputes. For that situation, it is customary (if illogical) to assume the alleged father has a 50% probability of being the father before DNA evidence is collected. The prior odds are then one, and the posterior odds are the same as the likelihood ratio, which is known as the 'paternity index'. In criminal trials, the problem of assigning odds can be avoided by presenting only the likelihood ratio, and the forensic scientist can testify that:

The DNA evidence (G) is LR times more likely if the defendant is the source of the crime sample (event H) than if some other person is the source of the crime sample (event $\bar{H}$).

DNA likelihood ratios

Calculation of the likelihood ratio is clarified by considering the nature of the DNA evidence G in some more detail. The evidence can be considered to be the events that profiles G_C of the crime sample and G_S of the suspect are both of type $\mathcal{A}$:

$$G : G_C = \mathcal{A} \text{ and } G_S = \mathcal{A}.$$

From the third law of probability:

$$\begin{aligned}\Pr(G|H,I) &= \Pr(G_C = \mathcal{A} \text{ and } G_S = \mathcal{A}|H,I) \\ &= \Pr(G_C = \mathcal{A}|G_S = \mathcal{A},H,I)\Pr(G_S = \mathcal{A}|H,I) \\ \Pr(G|\bar{H},I) &= \Pr(G_C = \mathcal{A}|G_S = \mathcal{A},\bar{H},I)\Pr(G_S = \mathcal{A}|\bar{H},I)\end{aligned}$$

So the likelihood ratio is:

$$\text{LR} = \frac{\Pr(G_C = \mathcal{A}|G_S = \mathcal{A},H,I)}{\Pr(G_C = \mathcal{A}|G_S = \mathcal{A},\bar{H},I)} \times \frac{\Pr(G_S = \mathcal{A}|H,I)}{\Pr(G_S = \mathcal{A}|\bar{H},I)}$$

Now the DNA profile of the suspect was determined at the moment of conception of that person, and cannot be influenced by whether or not the person left the crime sample, so $\Pr(G_S = \mathcal{A}|H, I) = \Pr(G_S = \mathcal{A}|\bar{H}, I) = \Pr(G_S = \mathcal{A}|I)$. The likelihood ratio reduces to:

$$\text{LR} = \frac{\Pr(G_C = \mathcal{A}|G_S = \mathcal{A},H,I)}{\Pr(G_C = \mathcal{A}|G_S = \mathcal{A},\bar{H},I)}$$

Further simplification follows from assuming that the DNA typing system is sufficiently reliable that two samples from the same person will be found to match. When the suspect is the perpetrator (event H), the crime sample must therefore be of type $\mathcal{A}$ if it is known that the suspect is of type $\mathcal{A}$, so $\Pr(G_C = \mathcal{A}|G_S = \mathcal{A}, H, I) = 1$, and:

$$\text{LR} = \frac{1}{\Pr(G_C = \mathcal{A}|G_S = \mathcal{A},\bar{H},I)}$$

A final simplification follows from assuming that the DNA profiles from two different people (the perpetrator and the suspect when $\bar{H}$ is true) are independent, so $\Pr(G_C = \mathcal{A}|G_S = \mathcal{A}, \bar{H}, I) = \Pr(G_C = \mathcal{A}|I)$, and:

$$\text{LR} = \frac{1}{\Pr(G_C = \mathcal{A}|I)}$$

This last equation is the conventional way of presenting DNA evidence. Only the probability with which an unknown person would have the profile is needed. The information I may contain information about the population to which that person belongs. A profile probability of 1 in a million can be regarded as giving

a likelihood ratio of a million. The evidence would be a million times more likely if the suspect is the perpetrator than if some other person is the perpetrator.

It needs to be stressed that the results in this section apply only to the situation where the DNA evidence refers to material left at the crime scene by the perpetrator, and there is no DNA evidence at the scene that does not provide a match to the profile of the suspect. If the evidence refers to a bloodstain found on the clothing of the suspect, for example, and the stain has a DNA profile matching that of the victim, then additional factors need to be considered: what is the probability that the victim's blood would be transferred during the crime? what is the probability that the suspect would have nonself blood on his or her clothing? what is the probability that nonself blood on the suspect's clothing would match that of the victim?

Nature of DNA profiles

To quantify the evidentiary strength of a matching DNA profile, it is necessary to decompose the profile into its component parts. DNA profiling examines the two 'alleles' a person receives, one from each parent, at a number of 'loci'. There may be several possible allelic types at each locus, and a person may have two copies of the same type and so be homozygous at a locus, or may have two different alleles and so be heterozygous. As an example, consider the collection of five loci marketed as AmpliType^R or 'Polymarker'. The loci are known as LDLR, GYPA, HBGG, D7S8 and Gc. Loci LDLR, GYPA and D7S8 each have two alleles, A and B, whereas loci HBGG and Gc have three alleles, A, B and C. A person with Polymarker profile $\mathcal{A} = AB, BB, BC, AB, BC$ is homozygous at locus GYPA and heterozygous at each of the four other loci.

For locus LDLR, the population proportion for the AB genotype is estimated as $2p_{A_{LDLR}}p_{B_{LDLR}}$, where $p_{A_{LDLR}}, p_{B_{LDLR}}$ are the frequencies of LDLR alleles A, B. For locus GYPA, the proportion for the BB genotype is $p^2_{B_{GYPA}}$, where $p_{B_{GYPA}}$ is the frequency of GYPA B alleles. At each locus, these products estimate the probability with which a random person in the population would have that 'genotype'. It has been assumed that the two alleles received by an individual at one locus are independent. To combine information over loci, it is further assumed that alleles at different loci are independent and the products for each locus are multiplied together. For the whole profile in this example, the 'product rule' calculation gives:

$$\Pr(G_C = \mathcal{A}|I) = 16p_{A_{LDLR}}p_{B_{LDLR}}p^2_{B_{GYPA}}p_{B_{HBGG}}p_{C_{HBGG}} p_{A_{D7S8}}p_{B_{D7S8}}p_{B_{G_c}}p_{C_{G_c}}$$

Dependent profiles

For a crime in which it is known that the perpetrator must have a DNA profile of type A because the evidentiary sample had that type, it is natural to consider the probability with which a random person in the population would have that type. This is the quantity $\Pr(G_C = \mathcal{A}|I)$ given at the end of the previous section. However, the evidential value of the sample needs to take into account the fact that there is a person (the defendant) who has already been seen to have that profile. The quantity of forensic interest is the conditional probability $\Pr(G_C = \mathcal{A}|G_S = \mathcal{A}, \bar{H}, I)$, and this can be quite different from $\Pr(G_C = \mathcal{A}|I)$.

Effect of relatives The largest effect of dependencies between the DNA profiles of two people is when they are related. Relatives have the possibility of receiving the same genetic material from their common ancestors and therefore having the same DNA profile. Some common values for the conditional probabilities are shown in **Table 1**, where p_i is the

Table 1 Effects of family relatedness

Genotype A	*Relationship*	$\Pr(G_C = \mathcal{A}\|G_S = \mathcal{A}, \bar{H}, I)$	*LR*
A_iA_j	Full sibs	$(1 + p_i + p_j + 2p_ip_j)/4$	3.3
	Parent and child	$(p_i + p_j)/2$	10.0
	Half sibs	$(p_i + p_j + 4p_ip_j)/4$	16.7
	Uncle and nephew[a]	$(p_i + p_j + 4p_ip_j)/4$	16.7
	First cousins	$(p_i + p_j + 12p_ip_j)/8$	25.0
	Unrelated	$2p_ip_j$	50.0
A_iA_i	Full sibs	$(1 + p_i)^2/4$	3.3
	Parent and child	p_i	10.0
	Half sibs	$p_i(1 + p_i)/2$	18.2
	Uncle and nephew[a]	$p_i(1 + p_i)/2$	18.2
	First cousins	$p_i(1 + 3p_i)/4$	30.8
	Unrelated	p_i^2	100.0

[a] Or uncle and niece, aunt and nephew, aunt and niece.

Table 2 Effects of population structure

		$\theta = 0$	$\theta = 0.001$	$\theta = 0.01$	$\theta = 0.03$
p = 0.01	Heterozygote	5000	4152	1295	346
	Homozygote	10000	6439	863	157
p = 0.05	*Heterozygote*	200	193	145	89
	Homozygote	400	364	186	73
p = 0.10	Heterozygote	50	49	43	34
	Homozygote	100	96	67	37

population frequency of allele A_i. To put the effect into perspective, numerical values of LR= $1/\Pr(G_C=\mathcal{A}|G_S=\mathcal{A}, \bar{H}, I)$ are shown for the situation when all alleles have a frequency of 0.1. The effects will be less for larger frequencies.

Effect of population structure If a locus has five alleles, there are 15 possible pairs of alleles, or genotypes, at the locus. For 10 such loci, the number of possible genotypes is 576 billion. The number of possible profiles greatly exceeds the size of the world population, but the probability of any particular profile is greater than the simple product rule calculation after that profile has already been seen once. This dependency is imposed by evolution. Individuals C and S may share a DNA profile simply because they belong to a finite population. Although they may not be in the same family, they each have (at most) 2^n ancestors n generations ago, and this number quickly exceeds the population size. Any two people have some ancestors in common and this leads to a low probability of shared profiles. This probability is greater for smaller populations.

The theory currently used by forensic scientists to accommodate this evolutionary perspective rests on the following argument. If the allele frequencies in a subpopulation are known, the probability of an unknown person C having a certain DNA profile is estimated by the product rule using those allele frequencies. This calculation does not use the knowledge that S has already been seen to have the profile. However, the usual situation is that allele frequencies are not available for the subpopulation. This may be for practical difficulties in taking a sample from that group, or it may be because the subpopulation is not well defined. Instead, allele frequencies are available (or can be estimated) for the whole population. Taking an average over subpopulations of the conditional probability $\Pr(G_C = \mathcal{A}|G_S = \mathcal{A}, \bar{H}, I)$ leads to the following single-locus results for homozygotes A_iA_i or heterozygotes A_iA_j:

$$\Pr(G_C = A_iA_i|G_S = A_iA_i, \overline{H}, I) = \frac{[2\theta + (1-\theta)p_i][3\theta + (1-\theta)p_i]}{(1+\theta)(1+2\theta)}$$

$$\Pr(G_C = A_iA_j|G_S = A_iA_j, \overline{H}, I) = \frac{2[\theta + (1-\theta)p_i][\theta + (1-\theta)p_j]}{(1+\theta)(1+2\theta)}$$

In these equations θ is a measure of population structure. It can be considered as a measure of variability of allele frequencies over subpopulations: the variance of the frequency A_i over subpopulations is $\theta p_i(1-p_i)$. It is also a measure of the relatedness of two alleles in the same subpopulation. Some numerical consequences of allowing for population structure are shown in **Table 2**, where all alleles have the same frequencies.

One of the difficulties with this theory is in assigning a value to θ. If observations were available from a series of subpopulations within a population, it is possible to estimate $(\theta-\phi)/(1-\phi)$, where ϕ is the probability that two alleles in the whole population have a single ancestral allele. The quantity ϕ is often assumed to be zero, so that it is θ being estimated from subpopulation data. However, it is the lack of subpopulation data that has made the 'θ-formulation' necessary in the first place, and the usual procedure is to assign a value, such as 0.03, to θ. It needs to be stressed that the equations apply on average for any subpopulation within the population.

Conclusion

The interpretation of matching DNA profiles often relies on some statements of probability. These statements refer to the chance of an unknown person having the profile given that a known person has

been seen to have the profile. A simple approach is to ignore the conditioning on the known person and use an estimate of the population frequency of the profile. However, this ignores the possibility that the known and unknown persons are related either by being in the same family or by being in the same population.

See also: **Evidence**: Statistical Interpretation of Evidence/Bayesian Analysis.

Further Reading

Aitken CGG (1995) *Statistics and the Evaluation of Evidence for Forensic Scientists*. New York: Wiley.

Evett IW and Weir BS (1998) *Interpreting DNA Evidence: Statistical Genetics for Forensic Science*. Sunderland, MA: Sinauer.

Faigman DL, Kaye DH, Saks MJ and Sanders J (1997) *Modern Scientific Evidence: The Law and Science of Expert Testimony*. St Paul, MN: West.

Robertson B and Vignaux GA (1995) *Interpreting Evidence: Evaluating Forensic Science in the Courtroom*. Chichester: Wiley.

Royall R (1997) *Statistical Evidence: A Likelihood Paradigm*. London: Chapman and Hall.

Schum DA (1994) *Evidential Foundations of Probabilistic Reasoning*. New York: Wiley.

Weir BS (1996) *Genetic Data Analysis II*. Sunderland, MA: Sinauer.

DETECTION OF DECEPTION

E Elaad, Division of Identification and Forensic Science, Israel National Police Headquarters, Jerusalem, Israel

Copywright © 2000 Academic Press

doi:10.1006/rwfs.2000.0466

Introduction

Lying in humans is commonplace and relevant to many areas in life. The successful implementation of a deceit may provide the deceiver with obvious advantages, and unmasking a lie may provide social and economic advantages for the lie catcher. This is especially true in criminal investigations where evidence is gathered from information elicited from suspects. Better understanding of the lying behavior may therefore be useful in finding out whether a given person is deceptive regarding a specific, usually criminal, event. Toward this end, efforts are invested in developing individual skills to detect deception from other people's verbal and nonverbal behavior. However, research in social and cognitive psychology has demonstrated the inability of human lie catchers to detect deceit. The experimental literature reported that observers rarely surpass 60% accuracy when detecting deceit, where 50% accuracy is expected by chance.

Furthermore, without using any instrumentation, even professional lie catchers such as judges, police interrogators, intelligence officers, psychiatrists, and polygraph experts are, on the average, only slightly better than chance at detecting deception.

Early Attempts to Detect Deception

People who were involved in the assessment of credibility looked for accurate detection methods. Many of the early attempts to detect deception were founded on magic and mysticism. Actually, all the ancient methods could be classified as trial by either torture, combat or ordeal. Torture was used to make the presumed guilty person confess his guilt. In ancient Greece torture was reserved for slaves and for strangers. Free citizens were not tortured. In Rome, torture was a widespread procedure that spared no one. The torture of freemen accused of treason or of other crimes against the state became an admitted principle of Roman law. Torture was used for the investigation of truth, not as punishment. Therefore, it was used for witnesses as well. When the testimony of slaves was required, it was necessarily accompanied by torture to confirm it.

However, there were some limitations to the use of torture. Women were spared torture during

pregnancy. Judges were not allowed to torture more than necessary, and it was only used when the accused was about to confess. Marcus Aurelius ordered the exemption of patricians and of the higher imperial officers. Diocletian forbade torture of soldiers, and after the adoption of Christianity, Theodosius directed that priests should not be subjected to torture.

The notorious Spanish inquisition used torture to detect the hidden crimes of those who were unfaithful to the church. In fact, the whole system of the inquisition was built on torture. The presumption of guilt led to the use of cruel methods to force the accused to confess. Hence, all the odds were turned against the accused and only few escaped.

Trial by combat was based on the adoption of divine judgment, and contestants let the outcomes of the battle decide who is truthful and who is not. Such a belief assumes that the adversaries should defend their claims themselves. However, in cases where the accused was unfit to fight, for example when a woman was accused, she was allowed to employ a champion to fight for her. At first, the champion was some member of the family. Later, it became the custom to substitute the contestant with a skilled champion, and professional champions sold their skill to the highest bidder. Champions had an interest not to inflict injuries, and they agreed on rules such as not to use teeth and hands in the fight. In medieval Italy, champions were recognized as a class with an established institution consisting of selected individuals. To enhance fairness, efforts were made to select champions who were equal in age, size and strength.

However, the efficacy of judicial combat is questionable. For example, in about the year 1100 Anselm stole the sacred vessels from the church of Leon. The merchant to whom he sold the vessels revealed his name to the church authorities. Anselm denied the accusation, offered to battle the merchant, and defeated him. Anselm was, therefore, proclaimed innocent. Later, Anselm confessed to his crime.

The injustice of both torture and trial by combat, their fallibility, and the advance of civilization encouraged the search for more peaceful modes to detect guilt. Torture and combat were abandoned.

The ordeal is another way by which people cast their doubts on a higher power. It was based on the belief that God will protect the innocent and punish the guilty person. For example, in 592 a bishop, who was accused of a crime, took an oath on the relics of St. Peter. It was evident that the accused exposed himself to immediate danger, if guilty. However, he performed the ceremony unharmed and this was a proof of his innocence. The literature refers to the ordeal of boiling water, the ordeal of cold water, the fire ordeal, the ordeal of balance, the ordeal of rice chewing, and a variety of other ordeals.

Unlike torture and combat, ordeals are presently practiced to detect deception. In Tahiti the priest digs a hole in the clay floor, fills it with water and stands over it praying to God. God is supposed to lead the spirit of the thief over the water and the priest, who constantly looks at the water will see it. It seems that the thief who stands with others in a circle is more anxious than others to see if the priest detected his spirit. He approaches the pool and the water reflects his image.

The ordeal of the red-hot iron is applied among Bedouin tribes. The suspect licks a gigantic duly heated spoon and if he does not burn his tongue he is acquitted. One would expect that many innocent suspects will thus burn their tongue. However, many suspects are proclaimed innocent after the test. This can be explained by the trust of the truthful suspect in the ordeal. The innocent suspect behaves normally, and the saliva on his tongue spares him the injury. The fear of the guilty suspect, on the other hand, reduces the activity of the salivary glands, and when the dry tongue touches the hot iron it is injured.

Similar to trial by combat, ordeals rely heavily on the belief of the suspect in God or in other mystical powers that control the outcome of the test. In an era of skepticism, ordeals can not provide the solution. Hence, it has to be replaced by methods that better reflect the spirit of the time.

The Myths of Hypnosis and Narcoanalysis

Gradually the detection of deception methods became more sophisticated and used more advanced technologies. Hence, it was thought that the use of hypnosis would assist in detecting deception. There are two major views to hypnosis. One believes that hypnosis represents a special form of consciousness which permits access to hidden parts of the mind. The other major view explains hypnosis by social psychological mechanisms and suggests that the hypnotized individual is affected by the social situation. Both views agree that hypnosis has no truth-compelling capacity. The person under hypnosis retains control and is able to judge events, and therefore can lie.

Another myth is that of narcoanalysis ('truth drugs'). Narcoanalysis was first used in psychiatric proceedings to facilitate communication with the emotionally disturbed patient. Drugs such as sodium amytal, and sodium pentothal induced relaxation, ease, confidence, and a marked verbal release. It seemed that the patient under the influence of the drug did not stop talking. The relief from inhibitions

and the decreased self-protective censorship of speech led to the idea that using drugs would reveal the hidden truth. However, it soon turned out that guilt-ridden people confessed under narcoanalysis to offenses they had imagined but had not committed, and others denied crimes that objective signs indicated they had committed.

Paper and Pencil Integrity Tests

Paper and pencil tests were developed to predict future thefts and other counterproductive behaviors of employees in the workplace. The tests are used in pre-employment screening and for other selection purposes. Some are overt integrity tests that ask applicants about their attitudes towards theft and other dishonest activities and about their own involvement in such behavior. Other tests disguise their purposes and include questions about dependability, conscientiousness, social conformity, trouble with authority and hostility. Integrity tests assume that honest people will be honest in all situations, and dishonest people are consistent in showing dishonest behavior.

Integrity tests are controversial. The controversy revolves around their effectiveness and the consequences of their use. Opponents claim that there is no such trait as honesty. The US Office of Technology Assessment (OTA) reviewed the research on integrity tests and concluded that there is no scientific support for the claim that integrity tests can predict dishonest behavior. The report asserts that integrity tests are biased against the applicant yielding a high rate of false positive errors (classifying an honest applicant as dishonest). These people will suffer from the stigma and be denied employment.

The American Psychological Association (APA) published another report which provided a generally favorable conclusion regarding the use of paper-and-pencil integrity tests in personnel selection. It was suggested that properly documented integrity tests can predict a number of dishonest behaviors, and their validity is comparable to many other tests used in pre-employment selection.

It seems that at present there is insufficient evidence to reach definite conclusions on the validity and the applicability of the integrity tests. However, there are indications that this direction may be promising.

Detection of Deception Through the Voice

Several devices were invented for the detection of emotional stress in the voice. The PSE (Psychological Stress Evaluator) has received more attention than others. The PSE was said to detect changes in the infrasonic frequency modulation that vary between 5 Hz and 20 Hz. These modulationss are controlled by the central nervous system and disappear during stress. The theoretical basis of the PSE was criticized as invalid, and experimental studies failed to establish its validity. The popularity of the PSE declined, and today it is rarely used. However, the interest in the voice as a potential channel for the detection of deception remains intact. New technologies were developed, and recently a computerized system called the 'TrusterPro' was introduced. At present its validity is unknown.

Statement Analysis

Other methods which were invented to detect deception are based on the analysis of statements. Transcripts of statements made by suspects or by witnesses in which they detail what they did or what they saw are analyzed. The methods assume that truthful statements differ from false ones in both content and quality.

One method, which was developed for the assessment of child witnesses in sexual abuse cases, is known as 'Statement Validity Analysis' (SVA). The SVA consists of two components: (1) a criteria-based content analysis (CBCA) in which 19 criteria have been proposed to reflect qualitative and quantitative differences between truthful and untruthful reports; and (2) examination of other evidence in the case. It is thought, for example, that a detailed report, which contains the exact description of the place, vivid description of people, and a step by step description of the events would suggest that the statement reflects the truth. Research conducted to validate the SVA yielded positive results for three CBCA criteria. It was found that deceptive accounts contained fewer details and were rated as less coherent than truthful reports. Dishonest suspects were also less likely to admit not remembering aspects of the event under consideration.

Another method through which deception may be detected is the Scientific Content Analysis (SCAN). The SCAN assumes that deceptive suspects will use more deviations in pronoun usage such as replacing I with you. Furthermore, deceptive suspects will present long introductions or omit the introduction altogether. Deceivers will also use many unnecessary connectors, such as 'after I left', and 'and then'. Intuitively, the SCAN may work but to date there is no scientific evidence on the validity of the SCAN.

A third method suggested that the number of distinct words in a statement (types) should be divided by the total number of words (tokens). It was sug-

gested that a higher type–token ratio may indicate deception. This is explained by the cautious approach of deceptive suspects who try not to reveal self-incriminating information. Therefore, they phrase their testimony with higher lexical diversity. Further research is required to support the validity of lexical diversity.

Psychophysiological Detection of Deception (Polygraph)

When we speak about detection of deception the first thing that comes to mind is the polygraph. The polygraph is a device that continuously measures and records physiological responses from an examinee who answers a series of questions. Recordings are made of respiration, palmar sweating and relative blood pressure. Changes in respiration are obtained from two pneumatic rubber tubes positioned around the thoracic area and abdomen. The palmar sweating, or electrodermal skin response, is obtained from stainless-steel electrodes attached to the volar side of the index and fourth fingers of the examinee's hand. Blood pressure is obtained from an inflated pressure cuff positioned around the upper portion of the examinee's contralateral arm (**Fig. 1**). Another measurement that is used less frequently is the peripheral blood flow obtained from a photoelectric plethysmograph placed on a finger.

After the test, the recordings are analyzed. The analysis emphasizes inhibition of respiration, electrodermal response amplitude and changes in the blood pressure and blood volume (**Fig. 2**).

The most common psychophysiological detection methods are the variety of procedures known as the Control Question Technique (CQT). Basically, the CQT contains three types of questions. (1) Relevant questions refer to the crime under investigation in the 'did you do it?' form (e.g. 'Did you break into Mr Jones's store last Saturday?'). Relevant questions are typically answered 'no'. (2) Control questions deal with undesirable acts in the past and pertain to matters similar to the crime being investigated but with a larger scope (e.g. 'between the ages of 12 and 16 have you ever taken something valuable without permission?'). The control questions are formulated during the pretest interview with the intention that the examinee will remain with some doubt about the veracity of his 'no' answer. (3) Irrelevant questions focus on completely neutral issues to which the affirmative answer is a known truth (e.g. 'are you sitting on a chair?'). Irrelevant questions are intended to absorb the initial orienting response evoked by any opening question, and to enable rest periods between

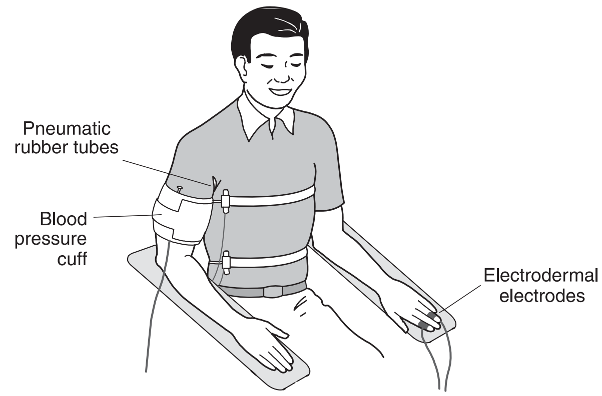


Figure 1 The polygraph attachments.

the more loaded questions. Typically, the whole question series is repeated three or four times.

The inference rule underlying the CQT is based on a comparison of the responses evoked by the relevant and control questions. Deceptive individuals are expected to show more pronounced responses to the relevant questions, whereas truthful individuals are expected to show larger responses to the control questions.

The CQT is available and convenient to use. Actually, all that is needed is a suspect who denies involvement with the crime and is willing to take the test. However, the CQT is also highly controversial. The controversy revolves around the plausibility of the rationale, the standardization of the test, the contamination by background information, the problem of countermeasures, and finally, the test's validity.

Actually, the vigorous controversy around the polygraph can be traced to the Frye case in 1923. Frye was charged with murder and denied the charge. He was tested with a systolic blood pressure test and was

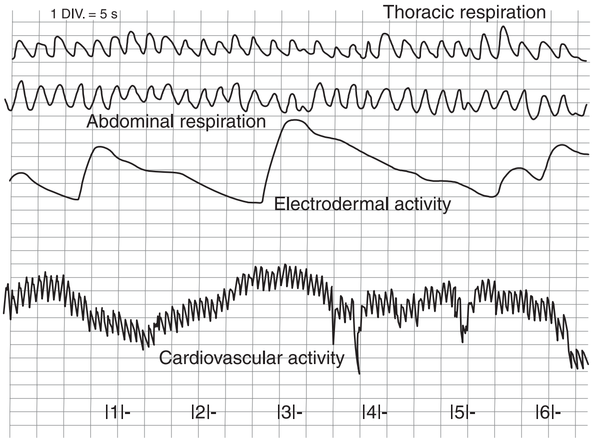


Figure 2 Illustration of respiration, electrodermal and cardiovascular patterns of recording on a chart.

found truthful. The court did not admit the test's results as evidence and in a landmark decision stated that: 'while courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs'. Since 1923 the polygraph changed. However, the Frye decision reinforced the reluctance of most courts to admit the polygraph results. In addition, the scientific community also ignored polygraph testing. Hence, practice has outpaced research and the polygraph was operated by practitioners who lacked training and expertise in psychology. This resulted in low standards and severe ethical problems, which further heated the controversy. Only recently the Daubert case overruled the Frye decision.

The Guilty Knowledge Test

The Guilty Knowledge Test (GKT), also known as the Concealed Knowledge Test (CKT), is a less controversial polygraph test. The GKT is used in applied settings to detect information that a suspect cannot or does not wish to reveal. The test utilizes a series of multiple-choice questions, each having one relevant alternative (e.g. a feature of the crime under investigation) and several neutral (control) alternatives, chosen so that an innocent suspect would not be able to discriminate them from the relevant alternative. The following is an example of a real GKT test which led to the apprehension of the culprits. At 22.45 h, four masked men entered the home/parish of a Nazareth priest. They gagged him and two other residents and locked them in a remote room after which they robbed the contents of the parish safe. Intelligence information led to the suspect who was a member of the priest's congregation. Twelve days after the robbery the suspect was brought in for a polygraph test. After denying knowledge of the correct answers, the suspect was asked the following questions: (a) 'What was the object used to gag the priest?' The following alternatives were presented to the suspect: Scarf, shirt, handkerchief, towel, sock, diaper, undershirt. The relevant item was a sock. (b) 'Which name was uttered by one of the robbers?' The following alternative answers were presented: Jacob, Salomon, David, Abraham, Samuel, Benjamin, Daniel. The relevant name was Abraham. It is assumed that only a guilty suspect will be able to single out and respond differentially to both the object that gagged the priest and the name the robber uttered. Innocent suspects, who have no guilty knowledge, are unable to distinguish crime-related information from other alternatives.

Inferences are made on the basis of the GKT by comparing the responses elicited by the relevant item with the responses to irrelevant items. Only if the responses to the relevant item are consistently larger, is guilty knowledge inferred. This provides a proper control against false positive outcomes, inasmuch as the likelihood that an innocent examinee might show consistently greater responsiveness to the correct alternative just by chance can be reduced to a low level by adding irrelevant items and by utilizing more GKT questions. In the robbery case, the guilty suspect, who later confessed, responded to both relevant items.

A recent review of 15 GKT mock crime experiments revealed that the rate of correct detection reported in simulated GKT experiments is quite impressive. It was found that across these 15 studies, 80.6% of 299 guilty examinees and 95.9% of 291 innocent examinees were correctly classified. Furthermore, in eleven of the studies, no false positives were observed. This supports the notion that the GKT can protect the innocent from false detection.

To establish the accuracy of the GKT in real-life criminal investigations, two field studies were designed. In both studies, the amplitude of the electrodermal response was used as an index of guilty knowledge. Excluding inconclusive outcomes, a very high detection rate for innocent suspects (97.9% and 97.4%, respectively) has been obtained. However, both studies reported a considerably lower detection rate for guilty suspects (50% and 53.3%, respectively).

The low detection rate obtained for guilty suspects, may be attributed to the differences between simulated and true examination conditions. First, the police investigator in the field cannot be sure that the guilty suspect noticed all the relevant details and remembered them at the time of the test. Therefore, an appropriate use of the GKT procedure requires the use of at least four or five GKT questions. In that case, the recognition of the other relevant items will compensate for overlooking one item by the guilty suspect. In any case, the small number of questions that were used in the field studies (mean number of 2.04 and 1.8 questions, respectively) may have contributed to the false negative error rate. Second, the use of a single electrodermal measure as the dependent variable undermines the test's accuracy. The second field study revealed that the addition of a respiration measure enhanced detection of guilty suspects to 75.8% while keeping the false-positive error rate relatively low (5.9%). Thus, a proper integration of two efficient measures increases the likelihood that the guilty suspect, who is responsive to at least one measure, will be detected.

An important issue related to polygraphic testing

concerns the extent to which a high accuracy level can be maintained when guilty suspects use specific point countermeasures in an effort to distort their physiological responses and yield favorable outcomes.

There are two major types of specific countermeasures: physical activities such as muscular movements, self-induced pain or controlled respiration, and mental activities such as thinking relaxing or exciting thoughts. The present evidence suggests that both countermeasure types may be efficient in distorting electrodermal responses, and it is possible for some deceptive individuals to beat the GKT in an experimental situation were the conditions are optimal and the participants receive information and training about the effective use of countermeasures.

However, in real life conditions, when the test is administered immediately after the interrogation, the suspect will not have the time to prepare, and the countermeasures may not be effective. Furthermore, the respiration measure is more resistant to the effects of specific countermeasures. This may suggest that more weight should be given to respiration measures in the GKT.

Another way to deal effectively with specific countermeasures is the use of slow-wave components of the event-related brain potentials (ERP). ERPs are measures of brain electric activity obtained from electrodes connected on the person's scalp. The P300 component of the ERP refers to a positive change in voltage of the peak which emerges at about 300 ms after presentation of a stimulus. The P300 signals special cognitive processing of recognized or infrequent stimulus. Therefore, it is ideal for detecting concealed knowledge. It has been shown that guilty participants produce P300 responses to stimuli recognized as guilty knowledge. The short latency of the P300 response is likely to undermine the effectiveness of specific countermeasures.

To conclude, the GKT is a standard psychological test. The questions are determined by the feature of the crime and may not depend on examiner or examinee factors. The experience in Japan, where the GKT is widely applied and the criminal justice system accepts its outcomes as evidence in court suggest that the GKT might meet the requirements of a legally accepted form of evidence.

Discussion

Polygraph tests were developed to help criminal investigators in situations where they are required to decide whether or not to believe the suspect. There are two major polygraph procedures, the CQT and the GKT. The more convenient and easy to operate CQT is widely used by law enforcement agencies. However, there is major controversy revolving around its rationale and inference rule, as well as around the empirical question of its validity. Opponents of the CQT suggest it should be replaced with the more standardized and scientifically based GKT. However, in many cases the GKT can not be an alternative for the CQT. Furthermore, GKTs are considered unimportant by federal law enforcement agencies in the US and are almost never used in criminal cases. A survey of FBI polygraph investigations estimated that the GKT might have been used in only 13.1% of them. The reason for this poor estimate is the difficulties involved in obtaining a sufficient number of proper GKT questions (i.e. questions that used items that can be identified by the guilty suspects but not by innocent people). Furthermore, the relevant information should be kept away from the public, the press, and the suspects during their interrogation. This can be achieved only by highly disciplined police investigators. In the United States most relevant case information is revealed either through the mass media or during the interrogation.

A contrasting view about the GKT prevails in Japan where the GKT is applied almost exclusively and its outcomes are accepted in court.

Effort should be invested in applying the GKT more frequently, but at the same time, the CQT should not be abandoned. The GKT can be used immediately after the initial interrogation as an elimination method. The CQT would be more useful later on after the suspicion against the suspect has been elaborated. The CQT is especially effective in cases where one suspect blames the other of committing an offense and the other claims to be falsely accused (e.g. complaints against police brutality and sex offenses). If both sides are being examined by two independent examiners and yield conflicting outcomes, even a modestly valid polygraph test would be sufficient.

Proper use of the various polygraph methods may facilitate police work and contribute to the welfare of innocent suspects. In many cases the polygraph is the only means by which innocent examinees can prove their innocence and free themselves from further interrogation.

See also: **Modus Operandi**. **Psychology and Psychiatry**: Overview; Psychiatry; Psychology.

Further Reading

Ben-Shakhar G and Furedy JJ (1990) *Theories and Applications in the Detection of Deception: A Psychophysiological and International Perspective*. New York: Springer-Verlag.

DePaulo BM (1992) Nonverbal behavior and self-presentation. *Psychological Bulletin* **111**:203–243.
Daubert v. *Merrell Dow Pharmaceuticals, Inc.* (1993) 113 S. Ct. 2786, 2799.
Ekman P (1985) *Telling Lies*. New York: W.W. Norton.
Ekman P and O'Sullivan M (1991) Who can catch a liar? *American Psychologist* **46**:913–920.
Elaad E (1990) Detection of guilty knowledge in real-life criminal investigation. *Journal of Applied Psychology* **75**:521–529.
Elaad E (1998) The challenge of the concealed knowledge polygraph test. *Expert Evidence* **6**:161–187.
Elaad E, Ginton A and Jungman N (1992) Detection measures in real-life criminal guilty knowledge tests. *Journal of Applied Psychology* **5**:757–767.
Farwell LA and Donchin E (1991) The truth will out: Interrogative polygraphy ('lie detection') with event-related brain potentials. *Psychophysiology* **28**:531–547.
Fry v. *United States*, 293 F. 1013, 1014 (C.C. App. 1923).
Gale A (1988) *Lies, Truth, and Science*. London: Sage.
Goldberg LR, Grenier JR, Guion RM, Sechrest LB and Wing H (1991) *Questionnaires Used in the Prediction of Trustworthiness in Preemployment Selection Decisions*. Washington DC: American Psychological Association.
Honts CR, Devitt MK, Winbush M and Kircher JC (1996) Mental and physical countermeasures reduce the accuracy of the concealed knowledge test. *Psychophysiology* **33**:84–92.
Honts CR and Perry MV (1992) Polygraph admissibility: Changes and challenges. *Law and Human Behavior* **16**:357–379.
Horvath F (1977) The effect of selected variables on interpretation of polygraph records. *Journal of Applied Psychology* **62**:127–136.
Lea HC (1971) *Superstition and Force*, 2nd edn. New York: Haskell House.
Lykken DT (1998) *A Tremor in the Blood: Uses and Abuses of the Lie Detector*, 2nd edn. New York: Plenum Trade.
Podlesny JA (1993) Is the guilty knowledge polygraph technique applicable in criminal investigations? A review of FBI case records. *Crime Laboratory Digest* **20**:57–61.
Porter S and Yuille JC (1996) The language of deceit: An investigation of the verbal clues to deception in the interrogation context. *Law and Human Behavior* **20**:443–458.
Raskin DC (1989) Polygraph techniques for the detection of deception. In: Raskin DC (ed.) *Psychological Methods in Criminal Investigation and Evidence*, pp. 247–296. New York: Springer-Verlag.
Reid JE and Inbau FE (1977) *Truth and Deception: The Polygraph ('Lie Detector') Technique*, 2nd edn. Baltimore: Williams & Wilkins.
Steller M and Koehnken G (1989) Criteria-based statement analysis. In: Raskin DC (ed.) *Psychological Methods in Criminal Investigation and Evidence*, pp. 217–246. New York: Springer-Verlag.
US Congress, Office of Technology Assessment (1983) *Scientific Validity of Polygraph Testing: A Research Review and Evaluation*. Washington DC: US Government Printing Office.
US Congress, Office of Technology Assessment (1990) *The Use of Integrity Tests for Pre-employment Screening*. (Report No. OTA-SET-442). Washington, DC: US Government Printing Office.

Digital Evidence *see* **Computer Crime**. **Electronic Communication and Information**. **Internet Crime**. **Photography and Digital Imaging**: Overview; Digital Imaging Enhancement.

DISASTER VICTIM IDENTIFICATION

J G Clement, Foundation Chair of Forensic Odontology, University of Melbourne, Australia

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0771

Introduction

A mass disaster is a catastrophic event of such magnitude or severity that it overwhelms the resources of the community directly affected and their capacity to cope or respond without external assistance:

> A national disaster is any natural phenomenon which causes such widespread human, material or environmental losses that the stricken community cannot recover without external assistance. (National Disaster Reduction Conference 1996).

There is almost always a substantial loss of life as well as damage to property and loss of crops and livestock.

It follows then that what may be a normal daily intake of bodies at a large, busy, urban municipal mortuary, could easily constitute a mass disaster in a small remote community. The landslide, collapse of ski lodges and deaths of 18 people at Thredbo in New South Wales, Australia in 1997 is a good example. One year after the event, the local community was still coming to terms with its loss, and the findings of the inquest and other investigations had still to be handed down.

At Port Arthur, Tasmania, in 1996 a lone gunman embarked upon a brief and savage murder spree, during which 35 people were shot dead and 22 people were wounded. The impact on the entire population of Tasmania was profound: 2 years after the murders, large sections of the community still struggled to cope with their legacy of psychological trauma. In this case, obvious criminality was involved and the perpetrator and his crime had to be investigated simultaneously with the identification of the victims.

In Zeebrugge, Belgium, a decade earlier, a large roll-on, roll-off ferry, the *Herald of Free Enterprise*, sank and almost 200 people perished. This constituted an 'open' disaster. Only a partial passenger manifest existed and so it will never be possible to be absolutely definite as to who was on board the ship at the time it capsized and sank.

In each of the three examples cited, and in almost every disaster, there is an urgent and pressing need to identify the victims on behalf of the next-of-kin. To be told that one's nearest and dearest has died, the body recovered and formally identified, is a personal tragedy. Ironically, not to know whether one's husband, mother, wife, son or daughter has been killed can be much worse if there lingers the likelihood that he or she might have been. This is often the case in open disasters, such as the earthquake at Kobe, Japan in 1996, the Tsunami tidal wave in Papua New Guinea in 1998 and most rail crashes, where the names and number of victims are not known with certainty.

In 'closed' disasters, such as aircraft crashes, involving a total loss of life for those on board, there usually exists a full crew and passenger list and the task of the investigators is really then to assign the correct identity to the appropriate remains. In open disasters there is the additional complication of trying to establish who was definitely a victim.

In all disasters, however, the families of the deceased need to be confident that the remains of their loved ones have been correctly identified and returned to them as quickly as possible. This allows grieving and funeral rites to take place – the first steps to be taken if people are to cope with their loss. In the case of the Thredbo disaster, some relatives felt strongly that a 'delay' of 3 days before bodies were released was unacceptable. Such feelings can often be allayed or assuaged if the next-of-kin are told of all the steps that have to be taken to insure that a formal identification is to be utterly reliable.

Little more than a century ago, Sir William Gladstone, a British Prime Minister, said:

> show me the manner in which a nation or a community cares for its dead and I will measure with mathematical exactness the tender sympathies of its people, their respect for the laws of the land and their loyalty to high ideals.

While his words seem somewhat quaint and stilted today, the principal message is abundantly clear: respect for the dead is an essential human trait. The manner in which we honor the dead is a reflection of our respect for the society in which we live. Forensic investigators are all too acutely aware of the responsibility placed upon them by the community they serve, but this only reinforces the mutual acknowledgement of the importance of their duties.

There are many legal reasons why it is necessary to identify the dead. In criminal charges of homicide or murder cases, it is an essential prerequisite to identify the victim even in the absence of the remains. In cases of sudden, unexpected or accidental death, the coroner, medical examiner or investigating magistrate has to establish the 'who', 'how', 'where' and 'when' of how the person died. It follows that in all mass disasters such legal obligations to investigate rest with the State. In cases where the nationals of one country die in another where customs, culture and legal process differ, this can lead to problems.

The inheritance of property, access to bank accounts, payment of pensions, compensation, and other very mundane but pressingly practical issues, depend for their settlement on the formal identification of the deceased, recorded in the form of a certificate of death.

Civil court action and litigation relating to issues of liability on (for example) the part of an airline contributing to a disaster needs the identity of the deceased to be formally corroborated. This may not only be the identity of the passenger who was someone's next-of-kin but perhaps the identity of the pilot of the aircraft at the time of the crash.

Issues relating to health and safety concerns in the workplace may require the driver of a long-distance freight truck to be identified. Commercial pressure in the workplace has been known to make employers coerce drivers into driving for too long without proper rest. This has led to fatal accidents. Similarly, pilots have been known to fly while under the influence of alcohol and other drugs. In all such cases, where guidelines, regulations and laws may have been brea-

ched, exactly who was at the wheel or the controls at the time of the accident has to be determined with the utmost rigor and certainty.

Principles of Identification

The identification of unknown human remains in the wake of mass disaster subscribes to the same principles that pertain for the identification of a single individual.

Identification is a comparative process. Preferably, the comparison is one with some form of dependable, tangible, pre-existing records that can be compared with a feature of the body remaining unchanged in death. Ideally, those features that are to be compared should be highly individualizing or unique to the individual; usually, the more detailed the antemortem record, the more likelihood there is of this being the case.

Sometimes, the amount of antemortem detail recorded far surpasses the original intent of the person making the record. For example, it is normal clinical practice in orthopedic surgery for radiographs ('X-rays') to be taken soon after an operation to replace a hip joint with a prosthesis. The surgeon may be interested in the position of the prosthesis, the interface between it and the bone, and other reasons driven by clinical need. However, the radiograph also records the shape, form and type of prosthesis (something already known by the surgeon and of no interest postoperatively). The radiograph also records a 'surgical signature'. Surgeons may use wire ligatures and fixations of a certain gauge; they may twist wires clockwise or anticlockwise, cut them short or leave them long. In this way surgeons are often able to recognize their own handiwork merely by looking at a postoperative radiograph of one of their patients.

While such a cursory examination may detect many familiar features, a detailed comparison of radiographs taken in the mortuary with one taken of a person in life can reveal such a constellation of 'points of concordance' that the identification of the deceased can be corroborated 'beyond any reasonable doubt'.

The ability to identify persons post mortem with such certainty is utterly dependent on the existence of reliable antemortem data. This may range from family photographs and medical and dental records to fingerprints and ungathered biological evidence.

The absence of any antemortem records usually implies that the investigator cannot gain access to the information, or the records are so scant or poor they are deemed unreliable, or the deceased is completely unknown. This frequently occurs in open disasters.

Where no individual records exist for comparison, the task of the identification expert is to narrow the range of possibilities as to what type of person the deceased had been. From skeletal evidence of teeth, was the deceased adult? Do the remains bear any racial traits or signs of ethnicity? These may range from naturally formed peculiarities of the dentition to self-inflicted tribal scarring or body art. Measurement of proportions of the body may be compared with pre-existing anthropological data to assign a putative ethnicity to remains.

The third very important step in the identification process draws upon the accumulated knowledge and experience of the investigator. There exists in the minds of the public the myth that, somewhere, all scientific knowledge is written down and can therefore be retrieved, if only the search is sufficiently diligent. This is not true. Much of what an expert knows has been acquired and absorbed almost unconsciously. Much of what is committed to textbooks can often be out of date because of the very nature of the publishing process. Real experts are cautious and prudent. They are reluctant to commit to writing much of what they have learned, not because they wish to be the 'keepers of some secret flame' but rather because, on specific issues, no opportunity has yet arisen for them to test or prove their insights in a legal context.

Where antemortem records are scant or of poor quality, the contribution that truly expert investigators can make to an investigation is incalculable. The process that they bring to the investigation is still a comparative one. It is not just a comparison of obvious, postmortem features with concrete, antemortem records: any competent scientist, medical or dental practitioner can do that. Their role may involve the expert gathering of data for comparison with type standards such as anthropological data. However, it really comes into its own and transcends the ordinary when the process is elevated to one of recognizing features that would normally escape detection by most observers, and having the knowledge and experience to appreciate the importance of those features. In this third case, the process is still a comparative one but it is a comparison between hidden features of the body with an incognizant expertise reinforced by a deep understanding of the rarer knowledge dispersed throughout the literature.

Methods of Identification

The most obvious means by which people may recognize one another is visually. Human beings are highly programmed to recognize one another in life. Body language strongly augments the spoken word. This has obvious value for social interaction, as the ability to communicate one's feelings, moods and emotions

is essential for survival. However, even the closest next-of-kin can find it very difficult to recognize members of their own family among a collection of other bodies arising from a mass disaster. This inability is frequently explained by outsiders as an example of 'denial' on the part of those struggling to cope with grief and loss. This is too conveniently simplistic. Witnesses to crime also find it hard to identify people from static 'mug shots'. Research subjects cannot always recognize that a series of different photographs may be all of the same person.

In the absence of pose, expression, gait, clothing and hairstyle, the human face becomes nondescript. Identification by visual means is very quick, cheap and convenient. Unfortunately for authorities under pressure to produce results, it can also be a very seductive option. Visual identification is subject to error by even the most well-intentioned identifier. It can also cause enormous psychological upset if the environment under which the process is attempted is inappropriate, or the feelings of the people involved are not fully considered. In the wake of the Lauda air crash in Thailand, pictures of extremely disfigured remains were put on public display in an attempt to reach those who might recognize the victims. In other societies, this might be seen as distasteful or shocking (**Fig. 1** and **Fig. 2**).

The final objection to visual identification relates to fraud. In the aftermath of the *Herald of Free Enterprise* shipwreck, people came forward and made fictitious claims relating to nonexistent relatives who were purported to be on board at the time of the loss of the vessel. Some went as far as making fraudulent insurance claims, for which a number were later tried and convicted. This is not an isolated example and, in the absence of strong corroboration by other means, visual identification alone (particularly by family and friends) should not be considered as sufficiently reliable to gain release of any body for burial, or, worse, from an investigator's point of view, cremation. Should a mistake be discovered later, it will be impossible to re-examine these bodies.

Victim clothing is sometimes used as an adjunct in the identification process. However, as travel increases, and with the increasing trend for mass-produced garments to be made in the less developed world, far from their eventual point of sale, little weight can be given to the clothing worn by victims of disaster. There are exceptions; for instance, in air crashes where some passengers may come from regions such as the Arabian peninsula, where traditional garb is still worn, it may be useful in the initial crude sorting of remains to be able to separate some passengers from others on the basis of dress. Similarly, the uniforms and insignia of the aircrew can be helpful

Figure 1 (see color plate 18) Inadequate body storage facilities. Aftermath of Lauda Air crash at Phu Khao Kao Chan near Bangkok, Thailand, in 1991. Bodies were stored outside with little protection from the heat or insects; decomposition was rapid. Body handlers were inadequately dressed, some with open footwear. Such a mortuary can become a secondary health hazard for the investigators. (Courtesy of Dr Carl KK Leung.)

as a first step towards their identification.

Jewelry is usually chosen carefully; it may be inscribed with the name of the giver and/or the recipient (or their nickname). It may also have been photographed for insurance purposes and, when found attached to a victim, can be excellent evidence for the corroboration of identity. Unfortunately much jewelry is not only precious but valuable as well. It can, and is, frequently pilfered or looted at the scene of a disaster (**Fig. 3**). A further disadvantage arises because it is attached to the periphery or integument of the body where it can be seen. Such parts are the most vulnerable to damage and destruction by trauma, fire, putrefaction or animal predation. This limits the application of the potential usefulness of jewelry for identification purposes.

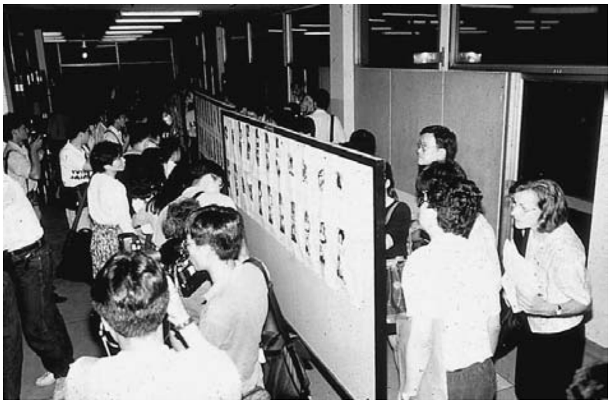


Figure 2 Lauda Air crash investigation. Photographs of the deceased put on display near to the mortuary to enlist the input of the general public to aid with identification. Many of the bodies were grossly disfigured. In many societies this would be seen as an inappropriate intrusion of privacy. (Courtesy of Dr Carl KK Leung.)

Documents such as passports and drivers' licenses can be very helpful circumstantial evidence of identity, in that documents carrying photographs of the bearer can be useful. However, documents can be stolen or transferred between people. A good example of this occurs when backpackers swap or sell airline tickets, which are then used in another person's name. This can lead to the situation where the passport of the deceased and the ticket are contradictory. Sometimes the only way to solve the conundrum is to locate the other person to exclude them from the investigation. This may be difficult, if not impossible, when the person concerned may have returned to their country of permanent residence.

Tattoos, scars and body art may be unique to an individual when the design and site of a tattoo or scar are both taken into account. Their vulnerability to destruction and relative rarity in the population at large limit their usefulness in disaster victim identification (**Fig. 4**).

Figure 4 Tattoo. When such evidence is present it can be invaluable for suggesting an identity for the deceased, who may have been photographed in life. However, tattooing is still relatively uncommon in the general population, which limits its availability for use in DVI. Furthermore, even when present ante mortem, tattoos may be destroyed post mortem by fire or putrefaction.

Fingerprints and footprints have very good evidentiary value in courts of law. Everyone normally has fingerprints, and they are considered for legal purposes to be practically unique. However, in many countries, only criminals have their fingerprints recorded on file ready for comparison. Should anyone from the rest of the population have to be identified, then latent prints taken from the victim's home or personal possessions have to be found, collected and transmitted. This can be a laborious process. The comparison of fingerprints still requires some subjective interpretation of the latent prints (which may vary in quality) and of the corresponding prints taken from a body that may have lain in water, or undergone desiccation or putrefaction, for some time, thereby changing their dimensions.

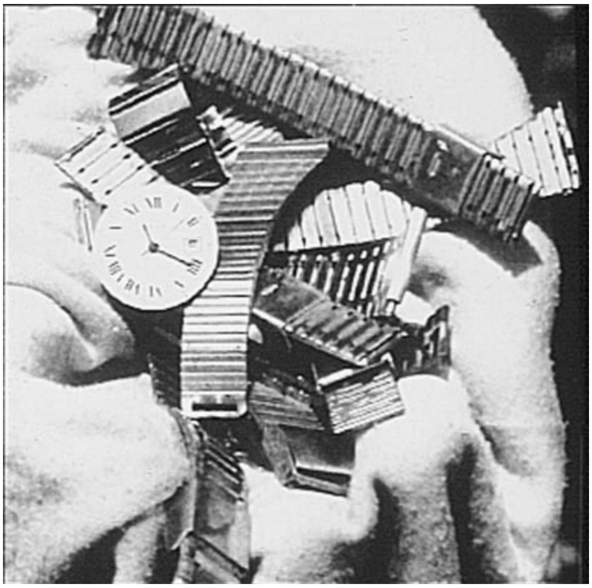


Figure 3 Lauda Air crash investigation. A selection of personal effects initially looted from the scene of the disaster and later recovered by the authorities. Even a temporary loss of evidence may seriously hamper investigations into the cause of the crash or the identification of the deceased. (Courtesy of Dr Carl KK Leung.)

Medical records often record transient events. A sore throat healed with the help of antibiotic therapy leaves no permanent scar or evidence that the illness ever occurred. This results in written clinical records but no corresponding postmortem evidence. Conversely, healed fractures, evidence of surgery, such as hysterectomy or appendectomy, or the insertion of pacemakers or metallic prostheses can provide very valuable evidence in corroborating identity: some is unique, some is not. Unfortunately, all such evidence is quite rare, which limits its applicability.

In contrast to medical records, dental records are almost ubiquitous in developed countries. They record treatments that are as permanent as orthopedic surgery but which have an additional advantage in that they can often be detected without dissection, together with features of the normal anatomy that are both individualizing and permanent.

With increasing emphasis on public health and fluoridation of many domestic water supplies, there has been a reduction in the prevalence of dental decay in the populations of developed countries. This does not necessarily reduce the number or quality of dental records. People who are concerned with dental health take steps to prevent damage to their front teeth while playing sports and have mouthguards constructed.

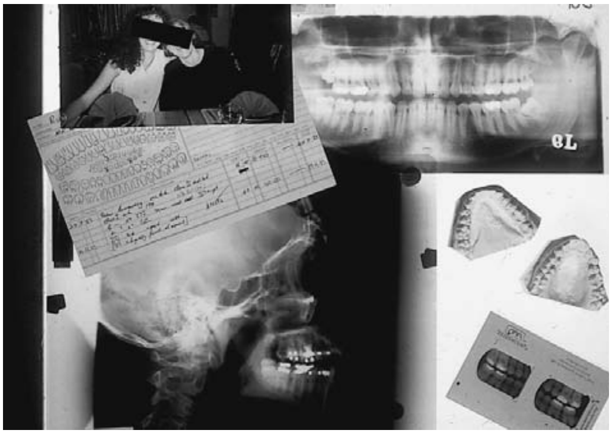


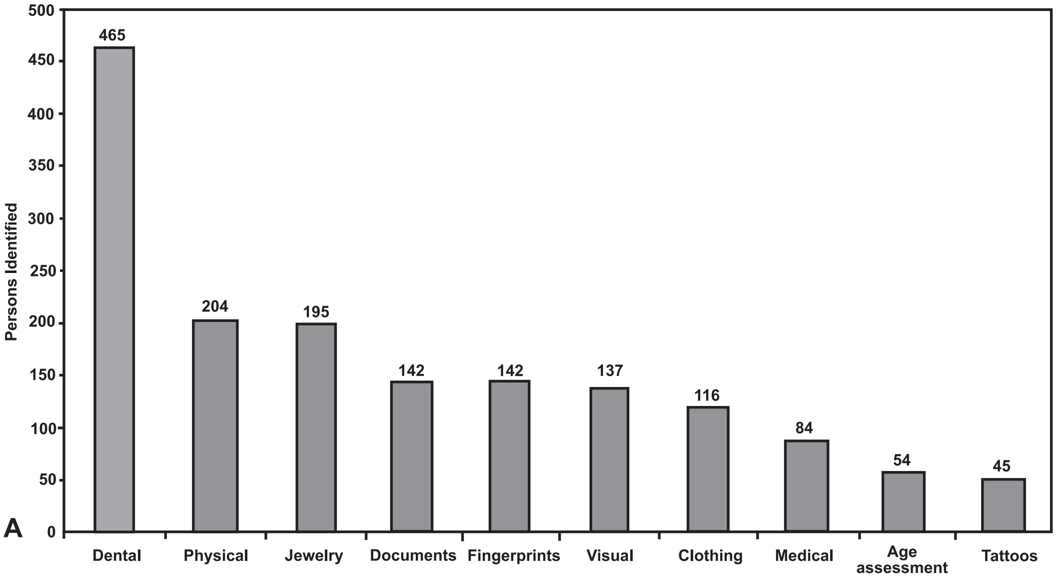
Figure 5 A selection of antemortem dental records in the form of written clinical records, accounts sent to the patient for payment, radiographs, plaster casts of the mouth and tooth crowns and photographs. This represents ample information in order to corroborate identity 'beyond reasonable doubt' and would suffice in a criminal investigation, which may follow from a mass disaster.

This requires dental impressions to be taken and plaster casts constructed, upon which the mouthguard is made. The casts record the size, shape and arrangement of the dentition and are an excellent source of dental records.

Others are concerned about the functionality or appearance of their teeth and seek the opinion of an orthodontist. As part of the assessment of the patient, plaster casts of upper and lower dental arches are constructed. The position of the teeth when they bite together is recorded. Radiographs of the head taken from the side record a silhouette of the soft tissues of the face, together with images of the teeth, jaws and skull. These are excellent for identification purposes, despite the victim never having received any dental treatment by means of fillings or restorations (**Fig. 5**).

The use of biological evidence for identification has come to mean DNA. Older methods establishing blood groups to narrow the possibilities for identification purposes are in decline and have been largely superseded. While the comparison of unique genetic material has obvious merit for corroborating identity, several basic problems arise in mass disasters. The first and most practical relates to fragmentation of bodies. In the Turkish Airlines crash in France in 1974, the 346 passengers and crew were reduced to more than 20 000 fragments. A similar fate befell the 105 people on board SilkAir MI 185 on 19 December 1997 when the plane crashed in Sumatra.

Selecting which pieces to identify is a problem. To attempt to identify all body parts requires access to a



B Identification by dental methods in five mass disasters.

	Cork	*Zeebrugge*	*Piper Alpha*	*Lockerbie*	*East Midlands*
Number of victims	329	193	167	270	44
No. of dental records received	279	137	132	252	38
Victims recovered (%)	40	100	81	89	100
ID of recovered bodies by dental records (%)	92	66	76	86	86

Figure 6 (A) Comparative usefulness of methods of identification in DVI. Data compiled from four recent disasters of different type and the identification criteria are those accepted by the investigating authorities. One of the disasters was a shipwreck, where many of the crew bore tattoos. Similarly 'visual identification' may be overrepresented and appear to be more reliable than is really the case. Identification by dental means has been demonstrated to be more than twice as useful as any other method in 'real-world' investigations. (**B**) Identification by dental methods in 5 recent disasters.
Cork = Air India AI 182 bombed near Cork in Eire 1985
Zeebrugge = *Herald of Free Enterprise* shipwreck near Zeebrugge 1987
Piper Alpha = Occidental Petroleum oil rig fire and explosion in North Sea, Europe 1988
Lockerbie = PanAm 103 bombed near Lockerbie, Scotland 1988
East Midlands = British Midland BD 092 air crash (engine failure) 1989 (Courtesy of Dr DH Clark.)

high level of expertise and huge laboratory resources. There are very few laboratories that have much inbuilt redundant capacity in terms of space, time or expert staff. In other words, to undertake an enormous study of comparing DNA, either with latent biological evidence or with DNA from blood relatives, requires other normally essential work to be deferred. This may not be possible.

Issues of cost are receding as DNA technology becomes more widely used, and so benefits from economies of scale, but the cost is still considerable and may still be prohibitive for those who have to bear that cost.

Another problem relates to the time taken to perform DNA analyses thoroughly. The extraction, purification and amplification of DNA is not always easy and takes time. If the body is putrefied, it may not be possible at all. What may be entirely feasible for one or two corpses can quickly become a practical impossibility in the time-frame available to the disaster victim identification (DVI) investigator.

The final hurdle for the use of DNA for DVI arises from skepticism or reluctance on the part of some courts to accept DNA evidence without question. Many of the barriers to its use have sprung from highly theoretical and somewhat spurious objections that have been spawned by the adversarial process relating to criminal cases. However, courts are now beginning to develop a more predictable and better informed position on the use of DNA for identification.

Selection of the Most Suitable Methods for Identification

From the outset of a disaster, it is important to understand who is in charge of the operation and its component parts. There is often widespread misunderstanding among the agencies concerned about chains of command and tiers of responsibility.

For those striving to identify the dead on behalf of their families and society at large (in that order), it is a wise precaution to assume that every investigation takes place in a potential crime scene. This has implications for the manner in which specimens are collected, documented and investigated, and a coordinated team approach after multidisciplinary conference is the only sensible way to go.

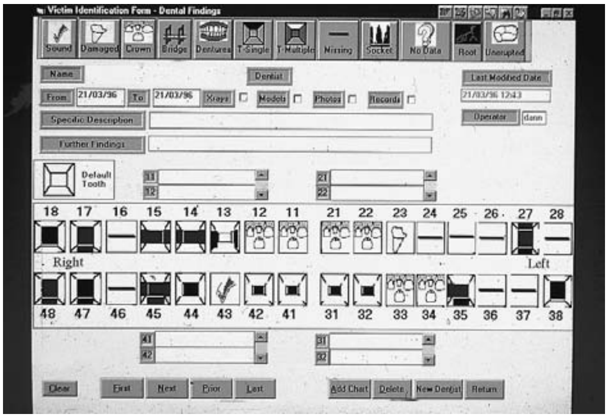
Jurisdictional problems always arise and it is important that all investigators understand what kind of evidence they have to provide in order for it to be admissible and thereby advance the investigation. There is no point in taking depositions of visual identification to the authorities if they require dental, fingerprint or biological evidence to prove identity. Similarly, it is pointless constructing an investigation plan that requires body parts to be exported to sophisticated laboratories overseas if the local investigating power insists on burying all the remains within 24 hours. In short, one has to be pragmatic and use the best methods available under each set of circumstances.

In recent studies it has been shown that about 80% of recovered bodies, eventually identified, will be identified by dental means; the usefulness of this technique far outstrips any other reliable method (**Fig. 6**).

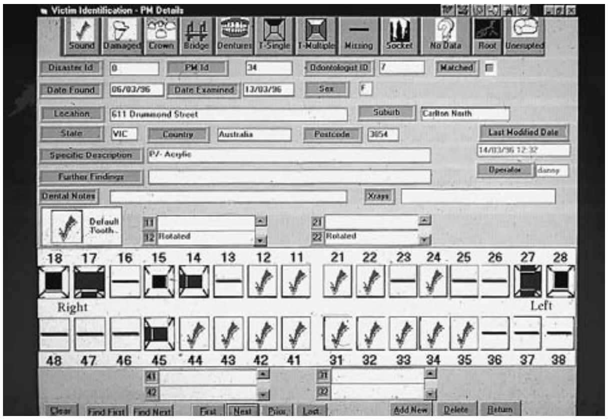
The selection of any method depends on the state of the bodies: fingerprints are useless when all skin has been burned away; dental methods are of very limited use when no antemortem records exist; and so on. For any method, or combination of methods, selected to suit the condition of the bodies, experts have to be available to conduct the work. In order for experts to function at their best, they have to be in a constant state of readiness. Somewhat contradictorily, this can only be achieved if they are normally engaged in other, closely related duties. For example, most forensic odontologists are dental practitioners who may also teach anatomy or pathology in university departments for most of their working lives.

The size of transportation disasters is on the increase and the loss of any large airliner with passengers and crew is likely to overwhelm the number of experts available, wherever the incident occurs. For this reason, contingency planning must make provision for the rapid importation of the required expertise.

In summary, the selection of the most suitable method, or methods, of identifying victims of disaster



A



B

Figure 7 DAVID computer program to handle dental data for use in DVI. DAVID is an acronym for Disaster And Victim IDentification. The program closely mimics the Interpol forms used in DVI. It has a Visual Basic front end and runs a Microsoft Access database in the background. It is therefore both fast and almost infinitely expandable without loss of performance. The program can use Federation Dentairé International and Universal notation. It handles both antemortem data and postmortem findings and rates their similarity. It can cope with partial remains and multiple antemortem records and can print final reports outside the mortuary in a clean environment. The program is very easy to use, having point and click techniques. DAVID is freely available from the World Wide Web. The site includes a manual to enable customization by the user. (**A**) Example of antemortem screen display for DAVID. The program can hold multiple antemortem dental records, each of which may be provided by a different practitioner for the same patient. Simple intuitive icons are provided to depict various treatments to the teeth. These can be altered by the user to suit themselves. The dialogue boxes above and below the odontogram are a compressed version of those found on the Interpol form and are scrollable. The background color of the form is yellow to match the corresponding Interpol form. Navigation is simple. All data entry is time and date stamped and only accessible via passwords. Only a single 'superuser' may amend or delete records. (**B**) Example of a postmortem screen display for DAVID. There is only one postmortem record permitted for each body or body part. Matching can be done in whole or in part. The background color of the form is pink to match the corresponding Interpol form. The default tooth can be set for any case to speed data entry when many teeth are in the same condition.

is a difficult and subtle task. The decision about which techniques to employ can only be made after an examination of the scene, a knowledge of what is required in a legal sense and an understanding of the resources available. In a world of finite resources, the cost, reliability and facilities required by each method have to be taken into consideration. The number and availability of suitably qualified experts, capable of working cooperatively, is something unknown until the event occurs. The time and resources taken to move bodies to a well-equipped facility need to be understood in the context of pressure from next-of-kin for prompt release of bodies and the need of the grieving to visit the disaster site.

Practical Considerations

Modern mortuaries are carefully designed to maximize the efficiency and performance of the staff who work in them. A temporary mortuary in a tent or aircraft hangar is a very poor substitute. Sometimes the environment where the incident has taken place is so inhospitable that there is no possibility of working there in a meaningful way. The Air New Zealand crash, at Mount Erebus in the snowy wastes of Antarctica, in 1979 and the Gulf Air crash, in the deserts of Abu Dhabi, in 1983 both required the removal of bodies from the scene to pre-existing mortuary complexes.

Where the environment is not so extreme the temptation arises to press into service some municipal building or large secure storage area. At the PanAm 103 crash in 1988, this resulted in bodies having to be carried up and down a curved staircase in the local town hall in Lockerbie. In the *Herald of Free Enterprise* disaster of 1987, bodies were initially stored in a gymnasium and then moved to a large warehouse inside a NATO base. In the first case, DVI investigators frequently had to withdraw from their tasks while next-of-kin went through the ordeal of looking through rows of bodies, searching for someone they had known. In the second situation, security could be assured but, in common with all temporary mortuaries, it suffered from problems of poor lighting, insufficient electrical supply, poor ventilation, poor drainage of body fluids, excess noise and poor infrastructure for communications. Temporary mortuaries

do not allow experts to do their best, and may actually put them at risk of injury or disease. Security of a crash scene or other disaster site is of vital importance. This is to prevent looting, the taking of souvenirs, interference by the media and animal predation; all may remove and destroy vital evidence. Sometimes disasters occur because of deliberate acts of terrorism; and, where bomb explosions are intended to kill and maim, it is common for secondary devices and other booby traps to be set to kill emergency service personnel. Other scenes may be mass graves filled with corpses during civil wars and genocidal acts. Again, for all investigators and their support teams, security at the site can be a cause for concern. Remains should be taken to a single, purpose-built mortuary, unless there are insurmountable reasons why this is not possible.

The inevitable delay in transporting remains will be repaid by the quality of information gleaned from the remains and the increased efficiency of operators. The ability to move bodies to a purpose-built mortuary may require the use of military personnel and resources, and presupposes that such a mortuary exists. Some of the larger, more modern mortuary complexes are able to expand body storage facilities in times of crises. This may include the facility to park refrigerated container trucks in a secure environment and to connect the refrigeration units to a mains supply. This frees the workplace of the noise and fumes from truck motors.

In the aftermath of the Abu Dhabi air crash at Mino Jebel Ali in 1983, bodies and body parts had to be taken to different mortuaries throughout the country. This meant the matching of badly fragmented bodies was made extremely difficult and put extra strain on communications.

Communications are always inadequate at the beginning of a disaster investigation. The telephone exchange at Lockerbie burned out through overload within hours of the crash of PanAm 103 in 1988. In the past, many emergency services have operated on frequencies peculiar to their own organization and have had to rely worldwide upon volunteer non-governmental organizations like WICEN (Wireless Institute Civil Emergency Network) to facilitate the exchange of mutually important operational information. Knowing how many bodies are at the scene, whether they are grouped or dispersed, ease of recovery, state of the remains and estimated time in transit to the mortuary – all assist in the preparations for the DVI investigator. For this reason it is important that someone with expertise in this area should be included in the small assessment team sent to the site of the disaster. Human resource estimations can be made and new team members recruited or stood down to match the circumstances.

In the rush to help, a normal-enough human reaction, several problems can arise. Firstly, people who are willing but not really able to cope can get involved to the point where they become casualties themselves. This impedes everyone else. Secondly, highly skilled people get involved in tasks outside their own area of expertise and dissipate their energy, so that when they really could be expected to work, they perform with reduced efficiency. Thirdly, experts push themselves too hard and for too long and start to make mistakes. Later they can feel guilty or angry with others who have to double-check their results. This can jeopardize the effectiveness of the team and hence the investigations.

For all of the above reasons it is imperative that the team leader has first-hand experience of DVI. In an operational sense, the team leader may not get too involved with mortuary teams, or the groups tracing, gathering and collating antemortem data, for whom he or she is responsible. It is important for all team members to know what is going on and what is expected of them. Such information cannot be given out on an individual basis and is better relayed at regular twice-daily briefings. All teams should informally debrief each other by sharing a good meal and refreshments at the end of each day. Investigators need to have privacy. They should not be housed in the same hotels as the press or the families of the deceased. Everyone needs to be alert to signs, in themselves and others, indicating poor coping strategies, and all team members should feel free to be able to walk away without explanation at any time.

Inter-agency rivalries are commonplace and can be extremely destructive. Negative criticism of other groups should be suppressed, and any suggestions relating to the performance or behavior of others should be channeled through team leaders.

Data generated by the investigation should be carried only along preagreed lines of reporting. No one, except the officially appointed spokesperson, should ever talk to the public, the press or the families.

All data should be entered onto paperwork specifically designed for the purpose. Interpol standards and documentation are widely adopted but others are used in certain jurisdictions. Computer programs have been designed to handle the vast amount of data generated in disaster investigation (**Fig. 7**). They vary in their usefulness and they should not be used unless the investigators have already used them, either in training or some other context, like a 'missing persons' database. It is courting disaster to try to train people in the finer points of a new computer program on what will be the worst day of their year. Similarly, it is essential that those people who have to fill in standard forms know exactly how to do it and

that all team members use the same abbreviations, symbols and shorthand.

Finally, when the antemortem record hunters/interpreters have gathered data, it is fed to teams charged with making comparisons with data gleaned post mortem. It is important for heads of specialist identification teams to pass information to each other on a regular basis. Interpol procedures include a large matrix in which all the findings of specialist groups can be seen together. This gives confidence in the final opinion passed to the coroner, medical examiner or investigating magistrate with whom the final decision about identity rests.

Conclusions

It is difficult to get most people to consider the unthinkable, yet unimaginable disasters will always occur. This inevitability of disaster should not give rise to a fatalistic attitude, for that removes the need to plan and prepare. Counterdisaster plans should be simple and up to date and understood by all with a need to know. A disaster plan should not be something 'done a few years ago', gathering dust on the shelf. There is the need for constantly revisiting plans to see how good they have been when implemented. This process of review, rehearsal and training should be continuous, so that the oral tradition is every bit as strong as the written word in the manual. In the words of the famous maxim in QANTAS Airline's Department of Safety and Environment:

> If you think safety is expensive, try having an accident.

Such costs are not merely financial. It behoves us all to mitigate accidents and loss of life by the use of good design, workplace practices and counterdisaster planning. However, in the end, we owe it to the families of those who perish to place the proper identification of the deceased at the top of our list of priorities when the unthinkable does, inevitably, occur.

See also: **Serology**: Blood Identification. **Deoxyribonucleic Acid**: Significance; Statistical Analysis. **Identification/Individualization**: Overview and Meaning of ID. **Odontology**. **Anthropology**: Skeletal Analysis.

Further Reading

Bowers MC and Bell GL (eds) (1995) *Manual of Forensic Odontology*. Colorado Springs: American Academy of Forensic Sciences.

Clark DH (ed) (1992) *Practical Forensic Odontology*. London: Wright.

Clement JG and Ranson DL (eds) (1998) *Craniofacial Identification in Forensic Medicine*. London: Arnold.

Cottone JA and Standish SM (eds) (1982) *Outline of Forensic Dentistry*. New York: Year Book.

Freckelton I and Selby H (eds) (1993) *Expert Evidence – Practice and Advocacy*, vol. 1. Australia: Law Book Company.

Hunger H and Leopold D (eds) (1978) *Identifikation*. Leipzig: Barth.

Plueckhahn VD and Cordner SM (eds) (1991) *Ethics, Legal Medicine and Forensic Pathology*. Melbourne: Melbourne University Press.

Stimson PG and Mertz C (eds) (1997) *Forensic Dentistry*. Boca Raton, FL: CRC Press.

Whittaker DK and MacDonald DG (eds) (1989) *A Colour Atlas of Forensic Dentistry*. London: Wolfe.

York LK and Hicks JW (eds) (1992) *Proceedings of an International Symposium on the Forensic Aspects of Mass Disasters and Crime Scene Reconstruction, FBI Academy Quantico*. Laboratory Division of the FBI, Washington, DC: US Government Printing Office.

Disaster-scene *see* **Crime-scene Investigation and Examination**: Major Incident Scene Management. **Disaster Victim Identification**.

DNA *see* **Deoxyribonucleic Acid**: Basic Principles; Databanks; Future Analytical Techniques; Mitochondrial Deoxyribonucleic Acid; Parentage Testing; Polymerase Chain Reaction; Polymerase Chain Reaction-Short Tandem Repeats; Restriction Fragment Length Polymorphism; Significance; Statistical Analysis. **Hair**: Deoxyribonucleic Acid Typing.

DOCUMENT ANALYSIS

Contents

Analytical Methods
Document Dating
Forgery/Counterfeits
Handwriting
Ink Analysis

Analytical Methods

V Aginsky, Riley, Welsh and Associates, Forensic Document Examinations, Inc., East Lansing, MI, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0477

Introduction

Tasks often requested for the analysis of documents involve distinguishing inks on the same or different documents and also identifying the source or date of a particular ink or paper. Usually these types of determinations are conducted with the help of analytical methods. There are many such methods available and the document examiner should know their capabilities and limitations.

Optical Examinations

Practically each examination of a questioned (*Q*) document starts with simple optical methods which allow observations in ultraviolet (UV), visible (natural daylight, filtered or unfiltered artificial light) or near infrared (IR) regions of the electromagnetic spectrum. These methods analyze color and luminescence of ink on paper, security fibers in paper, etc. They may also help in viewing obliterated writings, exposing alterations, erasures and substitutions, and discriminating between writing inks that appear to be of the same color.

The morphology of ink lines on paper is examined with the help of optical microscopy that uses visible light for illumination, and glass lenses for magnifying and focusing. This may allow classification of inks being examined, discrimination between inks being compared, or may, in rare occasions, individualize the writing instrument through its performance characteristics. In combination with spot color tests, this technique is also used for paper fiber analysis that allows the type of fibrous raw materials used for making the paper of *Q* documents to be determined.

Scanning electron microscopy is used when a highly magnified image (down to the nanometer range) of a micro-fragment of ink on paper or of paper itself is desired, for example, in cases when the sequence of crossing strokes is under examination.

Chemical Reactions

Spot chemical (color or crystal) tests are known to have been used for more than a hundred years for detecting both inorganic and organic ingredients of inks and paper. The spot or solubility tests are carried out both *in situ* (on the document itself) or on a removed sample. These tests are used to differentiate ink formulas, to presumptively identify the constituents of an ink formula, or to select a solvent suitable for the following extraction of the ink.

Spot color and solubility tests have been used for determining the sequence of crossing strokes of different inks and for evaluating the relative age of inks of the same formula and on the same paper.

Spectroscopic Techniques

Spectroscopic methods measure the absorption, emission, or scattering of electromagnetic radiation by atoms or molecules of compounds. The resulting spectra of the absorption, emission, or scattering of light are functions of wavelength and depend on the energy level structure of atoms or molecules. These spectra are useful for characterizing and identifying (e.g. with infrared spectra) compounds.

X-ray fluorescence is commonly used for solids in which secondary X-ray emission is generated by excitation of a sample with X-rays. The technique has found extensive applications in determining the

elemental profile of the ink and paper of suspect currency. This profile is then compared with the profile of genuine currency to uncover inconsistencies.

Energy dispersive X-ray microanalysis combined with scanning electron microscopy (SEM-EDX) is an important analytical method due to its ability to examine surface morphology with high resolution and depth of field, and to produce qualitative and quantitative elemental analyses of selected microareas by detection of characteristic X-rays. Many writing and printing inks contain detectable elements that can be used for characterization and comparison by SEM-EDX. In addition some writing inks have detectable trace rare organometallic compounds added which indicate the year of its production. Finally, the technique is a useful tool for the classification and differentiation of photocopier toners. In particular, it is useful in recognizing monocomponent process toners as they contain magnetic carrier materials (magnetite, ferrite) which are easily detected by SEM-EDX.

SEM-EDX has also been applied to the characterization of trace elemental profiles of pencils. The SEM-EDX analysis of both plain and coated photocopy paper has been used to provide a comparison, detect batch differences or to ensure that the contents of the minor inorganic components detected in the ink or toner samples cut out of the document, are not the result of their contact with the paper.

Other analytical methods that also have been used for determining the elemental composition of ink and paper are inductively coupled plasma mass spectrometry (ICP-MS) and neutron activation analysis (NAA).

Ultraviolet and visible absorption spectroscopy is used mostly for the analysis of organic materials. It measures the wavelength and intensity of absorption of near-ultraviolet and visible light by a sample.

UV–visible reflectance microspectrophotometry has been applied to measuring reflectance (absorbance) electronic spectra of ink on paper. The method allows discrimination between similarly colored inks at a considerably higher degree of certainty than it could be done using optical microscopy or evaluation by the unaided eye. Microspectrofluorimetry has been used for measuring the emission spectra of ink on paper and of some additives to paper (fluorescent fibers, optical brighteners).

Infrared spectroscopy measures the wavelength and intensity of the absorption of mid-infrared light by a sample. As the wavelengths of IR absorption bands are characteristic of specific types of chemical bonds, IR spectroscopy can be used to identify compounds. It should be stressed, however, that, if the components of interest are analyzed without isolating from the matrices, their chemical identification is practically impossible; as a rule, only characterization of the major functional groups of the compound can be accomplished. In order to produce conclusive identification, either peak-to-peak correlation using the spectrum of a known sample or a comprehensive software library of IR spectra would be required.

Fourier transform infrared (FT-IR) spectroscopy has been used for the characterization of organic components in many materials commonly examined during document analysis (ink, paper, photocopier toners, correcting fluids, etc.).

Diffuse reflectance infrared Fourier transform spectroscopy (DRIFTS) has been found to be a reliable, reproducible and selective technique for the classification and identification of photocopier toners. Compared with conventional dispersive IR spectroscopy, the DRIFTS technique provides spectra with a significantly improved signal-to-noise ratio, and therefore, it more effectively extracts data from toners that are normally highly absorbing in the infrared due to the large proportion of carbon black content.

Recently, FT-IR microspectrophotometry (a microscope attachment allows the infrared beam to be focused on an extremely small area) has been extensively used for the characterization and differentiation of writing inks and photocopier toners.

IR spectroscopy can also be used for document analysis in combination with other techniques. Thus, ink resin can undergo pyrolysis (see below), followed by IR analysis of the volatile gases generated. In most cases, spectra of the pyrolysis products resemble those of the parent substances. Even when they do not, the spectra are fairly reproducible; thus the reference spectrum of a known substance prepared in the same manner can be used for comparison with the material (ink, toner, paper) analyzed.

Raman spectroscopy (an emission technique in which a laser is directed onto the sample and a very small fraction of the scattered radiation displaced from the laser wavenumber by the vibrational wavenumbers of the sample, is measured) is used for the analysis of inks and photocopying toners in a manner similar to IR spectroscopy.

Chromatographic Techniques

Chromatography is a method used to separate, characterize and identify (e.g. with mass spectrometry) the components of a mixture. Since its introduction in 1903 chromatography has become a separation method that is now a widely accepted and recognized technique.

In document analysis, chromatographic techniques are extensively used for the characterization, comparison, source determination and dating of ink.

Paper chromatography

In paper chromatography the mixture to be separated is allowed to soak along the paper by capillary action; the cellulose in the paper acts as the adsorbent. The technique, as well as paper electrophoresis, has been used for differentiating ink samples.

Thin-layer chromatography (TLC)

This is a form of liquid chromatography that is used for separating nonvolatile organic and inorganic compounds. Among other analytical techniques applied to document analysis, TLC has been most extensively used both for discriminating inks and for identifying ink formulas (by comparison with a 'complete' set of standards, see below).

A typical procedure for the TLC analysis of ink is as follows. A sample of ink dissolved in an appropriate solvent, is deposited as a spot (or a band) on the starting line of a TLC plate that consists of a stationary phase immobilized on a glass, aluminum or plastic plate. The constituents of the sample can be identified by simultaneously running standards with the unknown. The bottom edge of the plate is placed in a reservoir with a solvent (mobile liquid phase); the solvent moves up the plate by capillary action. When the solvent front reaches a certain height (e.g. the other edge of the stationary phase), the plate is removed from the solvent reservoir. Inks are mixtures of many components, which move up the plate at different rates due to differences in their partitioning behavior between the mobile liquid phase and the stationary phase. Most separated components of inks are easily detected on the resulting chromatogram due to their own color. The other separated spots (colorless vehicle components) can be visualized with UV light or by treating the plate with an appropriate chromogenic or fluorogenic reagent.

Besides the characterization and differentiation of writing ink and the chemical identification of dyes removed from currency involved in a robbery and exposed to an exploding dye-pack, both conventional TLC and high performance thin-layer chromatography (HPTLC) have been applied to discriminating between stamp pad and typewriter ribbon inks, printing inks (including those used in counterfeit currency), photocopier toners containing dyes mixed with the carbon black pigment, jet printer inks and papers (tinting materials, optical brighteners, sizers and other components of paper can be separated and used to discriminate between paper samples). The method has been used both in its normal phase (hydrophilic stationary phase, e.g. silica gel) and reversed phase (hydrophobic stationary phase, e.g. RP-18 modified silica gel) versions, including gradient elution of ink samples by automated multiple development. Postchromatographic derivatization has been used for the visualization of separated chromatographic zones of colorless organic components of inks and paper. Scanning TLC densitometry has shown a high discriminating power with regard to inks that are indistinguishable to the eye having subtle differences in relative proportions of their dye components.

For over the past twenty years, different approaches using TLC have been used for determining the age of ink on documents. According to a so-called 'static' approach that deals with the analytical profiles of inks that do not change with age, the examiner determines the age or source of inks by using a collection of reference standards or by detecting tags, e.g. optical brighteners or other unique components specially added by the manufacturer. If the manufacturer of the ink analyzed is identified and its formula is shown to be unique (through a tag or unique formula known only by the manufacturer), the manufacturer's files are consulted to determine the initial production date of the ink. This allows one to establish whether a *Q* ink was available or not at the time the document was allegedly prepared. One obvious limitation here is that only a few inks actually contain unique dating tags.

Another ink dating approach measures the 'dynamic' characteristics of an aging ink, i.e. those that change with age. Several ink-dating techniques based on TLC, evaluate the age (date) of a *Q* entry relative to reference samples which are known dated entries written by ink of the same formula as the *Q* entry. These techniques primarily use TLC to identify a *Q* ink formula. However, it should be emphasized that, in fact, unless one is certain that the formula is proven to be unique (see above), the identification of the *Q* ink formulation with 100% certainty is hardly possible. The reason for this is that, on the one hand, inks of the same type and of similar color are very similar in their dye components (separated and detected by TLC) and, on the other hand, no matter how comprehensive the collection of reference samples is, it will never be complete. Hence, it follows that unless the formula is unique, there is always a possibility that a true match is not in the standard ink library.

This circumstance is of extreme importance and it should always be kept in mind when the examiner uses any ink dating technique that is based on the ink formula identification approach.

High performance liquid chromatography (HPLC)

HPLC is a form of liquid chromatography in which the stationary phase is packed in a separation column.

Components of a sample to be analyzed are separated by injecting a plug of the sample onto the column. These components pass through the column at different rates due to differences in their partitioning behavior between the mobile liquid phase and the stationary phase. The presence of analytes in the column effluent is detected by measuring a change in refractive index, UV–visible absorption at a set wavelength, fluorescence after excitation with a suitable wavelength, or electrochemical response. The separated analytes can also be identified with the help of a mass spectrometric detector.

HPLC has successfully been applied to the characterization and differentiation of ballpoint inks. It enables a high discrimination between inks having similar dye composition by separating and comparing their colorless components such as the ink vehicle components which are reliably detected in the UV region of the electromagnetic spectrum. If inks that are to be compared are on different papers, samples taken from the papers should be analyzed by the same procedure used for the ink-on-paper to ensure that sizers, optical brighteners, tinting materials and other chemicals that may present in the paper would not interfere with the analysis. The modern HPLC provides the examiner with a highly sensitive multiwavelength detection system (diode array detector) which will provide not only chromatographic profiles of the inks being compared but also the *in-situ* recorded UV and visible spectra of each eluting peak in the chromatogram. Obviously, such a combination of chromatographic and spectral data improves the ability of HPLC to discriminate between closely related inks. The ability of HPLC to discriminate between similar ink samples is also enhanced by increasing the resolving power with gradient elution.

HPLC has also been used for the analysis of nonballpoint pen inks, as well as printing inks (including those used in counterfeit currency), photocopier toners and paper.

Capillary electrophoresis (CE)

Performing electrophoresis in small-diameter capillaries allows the use of high electric fields resulting in very efficient separations. Due to electroosmotic flow, all sample components migrate in pH buffer towards the negative electrode. A small volume of sample (a few nanoliters) is injected at the positive end of the capillary and the separated components are detected near the negative end of the capillary. CE detection is similar to detection in HPLC, and includes absorbance, fluorescence, electrochemical and mass spectrometry.

Two versions of CE known to have been used for ink and paper analysis are capillary zone electrophoresis (CZE) and micellar electrokinetic capillary electrochromatography (MECC). In CZE separation is solely based on charge, but MECC enables separation of both charged and neutral or even hydrophobic molecules; it becomes possible by adding organic solvents and surfactants to the pH buffers.

CE has recently been applied to the analysis of ballpoint, roller ball, fineliner and marker pen inks, and has shown a very high resolving power that allows the efficient separation of both major and minor components of ink dyes (including their substitution derivatives and isomers) and, therefore, the discrimination between inks with similar dyes from different sources or different batches. The amount of ink-on-paper needed for the analysis is comparable to HPLC and TLC. To detect peaks on the ink electropheregram caused by the paper's constituents (optical brighteners, etc.), blank paper samples of similar size as those taken from the inked paper should also be analyzed.

Gas chromatography/mass spectrometry (GC/MS)

Gas chromatography (GC) is the most widely used analytical technique in forensic laboratories. The technique primarily involves the use of three components: an injector, a separation column (in a thermostated oven) and a detector. After vaporization in the heated injector, the sample is then transferred to the column through the use of a carrier gas. The individual sample components mix with the gas, travel through the column and are selectively retained by the stationary liquid phase contained within the column. Finally, a detector is utilized to produce a signal to a recording device. The resulting gas chromatogram is a series of peaks, each of which is characteristic of a particular substance.

It has been shown that the most selective GC determination of components of the complex mixtures can be achieved by the coupling of a micromass-spectrometer (mass selective detector) and capillary GC. Mass selective detector uses the difference in mass-to-charge ratio of ionized molecules to separate them from each other. Molecules have distinctive fragmentation patterns that provide structural information usually sufficient for identifying substances separated by GC.

Thus, gas chromatography/mass spectrometry (GC/MS) produces a mass spectral fingerprint for each sample component eluting from the column and, therefore, can allow discrimination between compounds having a very similar chromatographic behavior (close retention indices).

GC/MS has been used for the ink characterization,

batch origin determination and ink comparison. In the scan acquisition mode, the method allows identification of an ink's volatile solid ingredients among which can be nonreacted low molecular mono- or oligomers, reagents and also proprietary additives that are often contained in the resins, polymers or other components of ink vehicles (carriers). It has been shown that, even in old ink-on-paper, high boiling vehicle solvents can be detected and identified using the selected ion monitoring (SIM) acquisition mode; the detector is set to monitor ions specific to the solvents commonly used in the manufacture of inks.

Recently, the unique ability of GC/MS to efficiently separate ink volatile components and to quantify them at down to picogram level has been successfully used for developing ink dating techniques applicable to ballpoint, porous tip and roller pen inks, stamp pad inks, inks for jet printers, and other inks containing high-boiling vehicles.

Pyrolysis gas chromatography

GC is capable of separating volatile organic substances. Therefore, it is not directly applicable to the analysis of such nonvolatile substances as resins in inks or sizing materials in paper. However, pyrolysis of similar nonvolatile substances leads to their breakdown (thermal decomposition) into smaller compounds which are volatile enough to be analyzed by GC. A pyrolysis device is directly connected to the inlet of the gas chromatograph and the compounds produced by pyrolysis are separated and detected by the chromatographic system. The resulting pyrogram is a highly specific pattern of peaks which is a 'fingerprint' of the substance analyzed.

Pyrolysis GC with mass spectrometric detection (PyGC/MS) has been used for the characterization of nonvolatile organic components in inks and photocopier toners. The technique has provided high discrimination between closely related inks and toners.

See also: **Analytical Techniques**: Separation Techniques; Microscopy; Spectroscopy: Basic Principles; Presumptive Chemical Tests; Mass Spectrometry. **Document Analysis**: Forgery/Counterfeits; Ink Analysis; Document Dating.

Further Reading

Aginsky VN (1996) Dating and characterizing writing, stamp pad and jet printer inks by gas chromatography/mass spectrometry. *International Journal of Forensic Document Examiners* **2**:103–115.

Brunelle RL and Reed RW (1984) *Forensic Examination of Ink and Paper*. Springfield: Charles C Thomas.

Cantu AA (1995) A sketch of analytical methods for document dating. Part I. The static approach: determining age independent analytical profiles. *International Journal of Forensic Document Examiners* **1**:40–51.

Kirchner JG (1978) *Thin Layer Chromatography*, 2nd edn. New York: Wiley Interscience.

Lyter AH (1983) Analysis of writing inks by high performance liquid chromatography. In: Lurie IS and Wittwer JD (eds) *High Performance Liquid Chromatography in Forensic Chemistry*. New York: Marcel Deker.

Maehly A and Stromburg L (1981) *Chemical Criminalistics*. New York: Springer-Verlag.

Merrill RA, Bartick EG and Mazzella WD (1996) Studies of techniques for analysis of photocopy toners by IR. *Journal of Forensic Sciences* **41**:81–88.

ASTM (1992) Standard guide for test methods for forensic writing ink comparison. *1992 Annual Book of ASTM Standards*, Section 13, ASTM Designation: E 1422-91 Philadelphia: ASTM.

Tebbett IR (1991) Chromatographic analysis of inks for forensic science applications. *Forensic Science Review* **3**:72–82.

Tebbett I (1992) *Gas Chromatography*. New York, London, Toronto, Sydney, Tokyo, Singapore: Ellis Horwood.

Xu X, de Koeijer JA, de Moel JJM and Logtenberg H (1997) Ink analysis for forensic science application by micellar electrokinetic capillary chromatography with photodiode array detection. *International Journal of Forensic Document Examiners* **3**:240–260.

Zeichner A, Levin N, Klein A and Novoselsky Y (1988) Transmission and reflectance microspectrophotometry of inks. *Journal of Forensic Sciences* **33**:1171–1184.

Document Dating

D C Purdy, Forensic Document Examination Services Inc., Ottawa, Canada

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0483

Overview

The misrepresentation of dates on documents is not a recent challenge faced by forensic document examiners. In his book, *Questioned Documents*, Albert S. Osborn provided several examples of documents which were altered or backdated to make it appear as though they were written much earlier. Many frauds still involve document dating problems and forensic document examiners should diligently search for any clues that suggest a document was prepared some time other than indicated.

Various methods can be employed to backdate or fabricate documents. Such incidents can involve the relatively simple process of overwriting the date on a receipt to far more complex undertakings such as falsifying an entire document. Regardless of the method used, dating suspect documents is a very challenging problem for the document examiner and should be approached cautiously.

The most straightforward method for solving a dating problem considers the types of office equipment and technologies used to produce the questioned document. The date on a document can be proven false if the instruments and materials used to produce it were unavailable when it was supposedly prepared. A second method takes into account certain features in the contested document that vary over time. Defective letters produced by a worn typewriter or photocopier 'trash marks' originating from dirt on the platen glass are two examples of this type of evidence that has dating significance. A third method involves the analysis of materials that make up a suspect document. For example, some speciality papers or writing inks contain materials added to these stationery products to improve their quality. If it can be established that these materials were introduced to the market on a specific date, any document in which they are found to be present must have been prepared at a later time. A fourth method involves knowledge of the aging of several components such as ink or paper.

The following sections describe different areas that can be examined to determine when a document was drawn up or whether its date is false. The results of these tests do not always provide conclusive evidence of fraud. They can, however, draw attention to irregularities that must be reconciled before a suspect document can be relied on as a genuine.

Paper Products

Watermarks

Conventional paper watermarks are produced during the manufacturing process by a 'dandy roll' cylinder located at the beginning of the papermaking machine where paper is formed into a web. The dandy roll cylinder consists of a woven wire gauze onto which raised designs are soldered or otherwise attached. A watermark is created when the relief areas of the dandy roll press into and displace paper fibers.

Paper mills usually maintain accurate records concerning their watermarks. Once the paper manufacturer of a questioned document is known, the company can be contacted to determine the earliest date that a watermark design was used. Any document dated earlier than this time must have been backdated.

The design of watermarks can also change over time as relief areas of a dandy roll suffer damage through normal wear and tear. Detached or broken wires produce slight but visible changes in the design which is transferred to the paper. Paper mills usually keep records when dandy roll damage occurred and when repairs were made. This information can be very helpful in narrowing the period during which a watermarked paper was manufactured.

A few paper companies have intentionally changed the design of their watermarks from time to time. Such watermarks are said to contain a 'date tag', which will often indicate the year that a sheet of paper was produced. For example, Southworth Paper Company placed a short bar below the letters in their watermark to indicate the last digit of the year in which the paper was manufactured (**Fig. 1**). If a document bears a watermark that was not in existence when it was allegedly dated, the genuineness of its date must surely be challenged.

When using watermarks to date paper, it is strongly recommended that the paper manufacturer be contacted to verify the time period when the noted features were present.

Paper composition

Over the years, different fillers, surface coatings or chemical additives have been added during the paper making process to improve the quality of the product. Other changes in the manufacturing processes have occurred for economic or environmental reasons. These innovations and modifications can establish the earliest date or period a particular sheet of paper was manufactured.

Many North American paper manufacturers stopped producing acidic paper in favor of alkaline or neutral process papers during the late 1980s and early 1990s. A simple pH test can indicate if a questioned document was produced before its purported date. This finding can be corroborated if

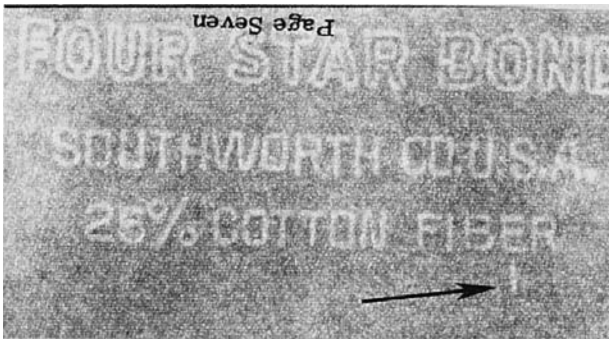


Figure 1 The short vertical bar beneath the letter 'E' in this watermark confirms the sheet of paper was manufactured during 1966.

certain chemicals that were introduced after the date on the document are present in the paper. For example, when mills converted their operations to an alkaline process, many also began using calcium carbonate ($CaCO_3$) as a substitute for titanium dioxide (TiO_2) in order to improve the brightness and opacity of papers. Caution should be exercised when interpreting such evidence and the paper manufacturer should be consulted to confirm when the observed processes and materials were introduced.

Speciality papers can also contain information of dating significance. For example, NCR (No Carbon Required) paper first appeared in the United States during 1954. The formula for manufacturing this product was changed several times during the 1960s and 1970s. In 1972, NCR developed a coding scheme to identify the source and date of its papers. Trace amounts of various high atomic weight elements have been added by other manufacturers as a means of tagging their products. The dates of documents produced on speciality papers that contain tags can be verified by taking such information into account.

Envelopes

Envelopes are often discarded once their contents are removed. This is unfortunate since an envelope may contain important information about when it was mailed and possibly when its contents were prepared. The following envelope areas can have dating significance: postage stamps, postage cancellation marks, envelope shape and printed information.

Postage stamps affixed to envelopes can be examined to determine if they were available when the envelope's contents were prepared. A new postage stamp is released for sale as a 'first day cover' on a particular date. Postal officials or a knowledgeable stamp collector should be able to provide the precise date a stamp was issued. Once this date is known, the envelope and its contents must have been mailed some time after this period.

Stamps on many envelopes bear cancellation marks that are applied by the post office. Even if a cancellation mark is not legible, the format of the mark, the way it was struck and the chemical composition of ink can serve to establish the period when it was applied.

Occasionally, logos or product codes are applied to envelopes while they are being manufactured which can have dating significance. The impression shown in **Fig. 2** was found on the inside flap of an envelope manufactured by Tenison Envelope Company. This mark represents the stock number (20), the initials of the company that placed the order (EC) and the number (5) corresponding to the last digit in the year it was manufactured. The envelope manufacturer should always be contacted to confirm the accuracy of dating information.

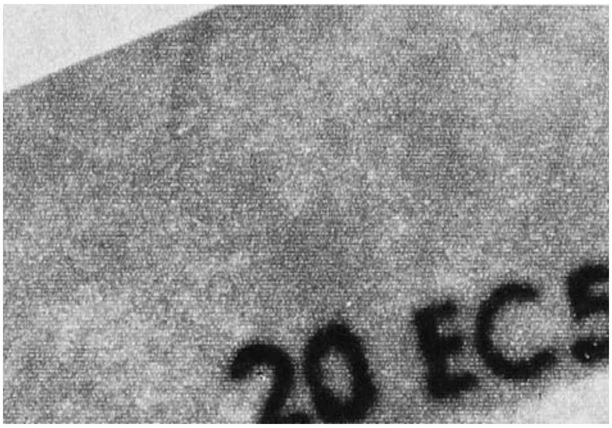


Figure 2 A notation '20EC5' on the inside flap of an envelope represents the stock number (20), the intials of the company that placed the order (EC) and the last digit in the year of manufacture (5).

The design and appearance of some envelopes are unique to their manufacturers and these features may well indicate when they were produced. These include, but are not limited to, the following:

- small irregularities along the edges of the paper related to a damaged die stamp;
- types of adhesives applied to the side-seams and flap areas of the envelope;
- striation patterns in adhesive layers associated with the method of application.

Other areas sometimes overlooked are addresses which appear on an envelope. A particular mailing or return address may not have existed when the document was supposed to have been sent. Postal or zip codes change from time to time and these should always be checked to insure they existed during the period in question.

Inks and Writing Instruments

One of the most challenging dating problems facing the document examiner is estimating when a particular document was signed or written. If a document was supposed to have been written many years ago, it may be possible to prove it was backdated if the type of pen and writing materials used were not available at that time. Important milestone events concerning the development of modern writing materials are shown in **Table 1** along with their dates of introduction.

Clues as to when a document was signed can also be found by analyzing the questioned writing ink. A small sample of ink removed from a document can be

Table 1 Significant dates of introduction in the history of writing instruments

Year	*Historical development*
624	Earliest reference to the quill pen
1662	Pencils made in Nuremberg, Germany
1700	Early reference to steel pens
1780	Steel pens made in England by Samuel Harrison
1857	First appearance of 'copying pencils'
1945	Ballpoint pen first marketed in New York city
1951	Felt-tip markers introduced
1955	Liquid lead pencil introduced
1963	Fiber-tip pen first produced
1968	Roller ball pen first produced
1979	Eraser Mate erasable pen introduced by Paper Mate

separated into its solid components by thin layer chromatography (TLC). The result of this analysis is a chromatogram that isolates the different dyes present in the ink formulation on a coated glass or plastic plate. Success of this method relies on the different physical and chemical properties of the ink and the existence of a sufficiently complete set of ink reference standards.

TLC can also detect the presence of tags which have been added to some writing inks by their manufacturers. During the 1970s, several US ink producers participated in an ink tagging program organized by the Alcohol, Tobacco and Firearms (ATF) Laboratory in the United States. This scheme urged ink manufacturers to add trace amounts of different materials with distinct properties to their inks. These materials would be changed annually and thereby indicate the year an ink was manufactured. By 1978, approximately 40% of writing inks produced in the United States contained such dating tags. Although this initiative greatly increased the ability of forensic scientists to date domestic writing inks, the continued growth of imported products threatened the success of the program. Although most ink manufacturers withdrew from the tagging program by the early 1980s, documents purportedly written before this period may contain chemical taggants that suggest they were manufactured at a much later date.

Ink chemists have observed that many writing inks begin to change or age the instant they are applied to paper. Most people have noticed writing inks fade or become lighter with the passage of time. In addition to this obvious physical transition, investigations have shown that the chemical composition of an ink also changes over several months or years. These effects are especially true with respect to the color, solubility and solvent volatility of the writing inks.

In 1995, Brunelle suggested a technique wherein two samples of a single ink entry are removed for testing. After artificially aging one sample by exposing it to heat, both samples are analyzed and the test results compared. It was discovered that ink parameters level off after several years of natural aging. This artificial aging technique is based on the hypothesis that induced aging will take an ink to where it would be if it had aged naturally. If differences between the two tested samples are slight, the result suggests the ink entry was on the paper for some time. Greater differences in solvent extraction are indicative of more recent ink entries.

Other testing methods rely on sophisticated analytical techniques such as gas chromatography/mass spectrometry (GC/MS) to measure the concentration of volatile components in an ink sample. This technique also requires two samples be taken from the suspect ink entry. After exposing one to heat, both samples are tested and the extent to which their solvent components differ provides an estimate of when the ink entry was written. This method is better suited for entries made within 12 months of testing.

The described methods are beyond all but a few specialists who possess the equipment, knowledge and experience needed to analyze and date writing inks. Some controversy still surrounds certain ink testing methods and further validation studies could resolve these debates.

Commercially Printed Documents

Many documents subjected to forensic examinations take the form of documents with letterheads, contracts, envelopes, notary records, receipts and other types of printed stationery. Apart from typewriting, handwriting and other information they may contain, commercial printing on documents can be used to establish whether they were produced during or after a certain period.

Minuscule printing defects such as irregular letter outlines, uneven inking or small breaks in line work can associate a questioned document with a particular stationery order produced by a commercial printer. Once the company that produced a printed document is identified, more precise information about when the order was delivered and the earliest time the stock was put into circulation can be determined. Access to samples from the order retained by the print shop can also be of value when attempting to date commercially printed documents.

A coded mark within the body of a print job can also provide important information about a print job. For example, the bottom of a medical form shown in **Fig. 3** bears the notation 'AC1215R0'. The printer used this number to trace the advertisement to the client. Information about the order indicated when

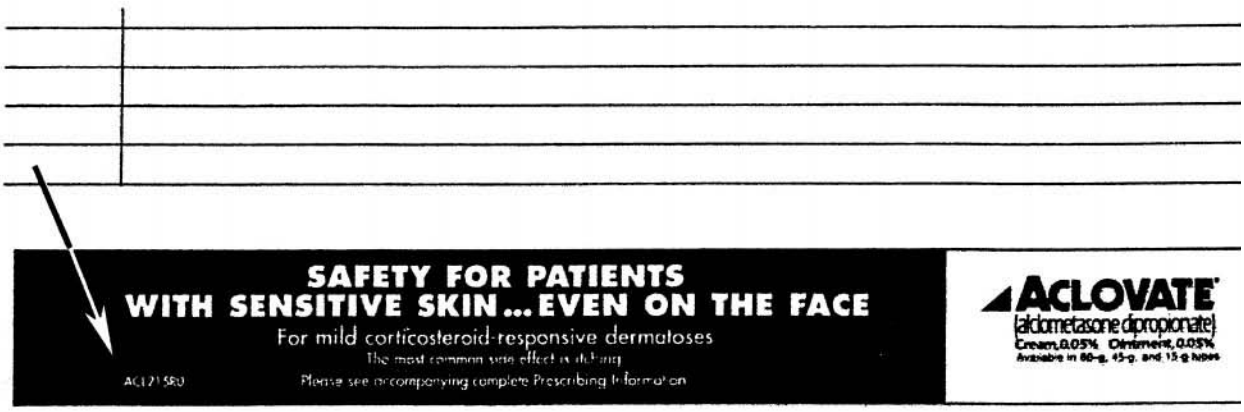


Figure 3 The alphanumeric code 'AC1215R0' within this advertisement on a patient's medical form was used to establish the date the stationery was printed.

the stationery was distributed. In this case, the advertising campaign was not authorized until three years after the patient records were supposed to have been written on the printed form. This provided irrefutable proof that the medical records had been prepared long after the patient was examined by the physician.

Typewriting

The typewriting technology used to produce a questioned document is one of the first factors that should be considered when its date is at issue. During the last century, many advances have occurred in the development of the modern typewriter. Some important events and when they occurred are listed in **Table 2**.

The date a typewritten document was prepared can be determined in other ways. One method considers the typestyle which appears on a questioned document. The shape and size of typed letters can indicate the make(s) and model(s) of typewriter(s) which might have been used to produce the typewriting. The results of searching a large collection of typewriter specimens can indicate that the questioned typestyle was introduced to the market on a particular date. Should the typestyle's date of introduction be later than the date on the suspect document, the questioned document must certainly be regarded with suspicion.

A second method of dating typescript takes into account any typeface defects present in the questioned typewritten text. Typewriters contain many moving parts which gradually become worn or defective with use. These defective components produce misaligned or damaged letters that become quite apparent when examined with a microscope. Subsequent adjustments or repairs by a service technician can create further changes to the appearance of typewriting produced by a machine. The dates when typewriter damage occurred or disappeared are very significant for dating purposes.

Table 2 Significant dates of introduction in the development of the typewriter

Year	*Technological development*
1909	First use of bi-colored ribbon (Underwood)
1927	First use of carbon ribbon (Hammond-Varityper)
1944	IBM Executive proportional spaced typewriter
1956	Remington Statesman the first proportional typewriter by Remington
1960	First Underwood proportional spaced typewriter
1960	Underwood electric standard typewriter with duplex carbon and fabric ribbons
1961	IBM Selectric I dual pitch single element typewriter
1963	First use of IBM Selectric polyethylene film ribbon
1971	IBM Selectric II dual escapement, half backspace machine
1971	Tech III ribbon cartridge for IBM Selectric
1972	First daisywheel produced by Diablo Systems
1973	IBM Correcting Selectric II with special lift-off ribbon
1975	Thermal transfer ribbon developed by IBM
1977	First use of polyurethane ribbons (Olivetti)
1978	First dot matrix printer for personal computer (Epson TX 80)
1982	IBM Electronic 65 and 85 typewriters with triple pitch and right justification
1982	Brother EP-20 seven-pin thermal typewriter
1984	Diablo releases EPM 1 – first thermal ribbon transfer printer
1984	IBM Quietwriter with nonimpact thermal print head
1984	Quietwriter ribbon by IBM

If a typewriter is not cleaned regularly, oil, ribbon particles, dirt and paper fibers can accumulate within the crevices of certain letters. When dirty typefaces strike the paper through the ribbon, the letters appear filled-in rather than clear letters and numbers. These imperfections will remain until the dirt is removed by cleaning the typefaces. Access to uncontested documents produced on the same typewriter over a period of time will reveal when changes to the appearance of the typescript occurred. **Fig. 4** shows how the appearance of typeface dirt and damage can expose a fraudulent document.

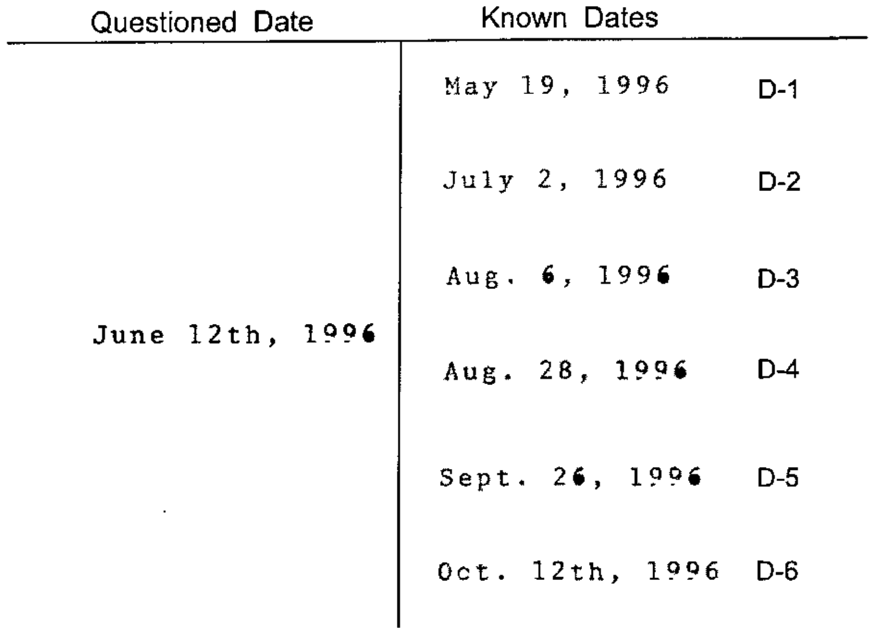


Figure 4 The questioned document could not have been typed on June 12th, 1996. Damage to the digit '9' and the filled-in body of the '6' occurred sometime after August 6, 1996.

Typewriter single-strike and correcting ribbons can also indicate the date when documents were produced on a particular typewriter. A used single-strike ribbon will contain impressions of all the characters struck by the machine in chronological order since the ribbon was last changed. If the typewriter ribbon used to produce a questioned document is available for inspection, it can be examined to insure the date of a questioned typewritten document is contemporaneous with the dates of typed documents which precede and follow it. If it is not, dated correspondence appearing immediately before and after the location of the question passage can serve to determine the approximate period when the contested document was typed.

Correction fluids applied to conceal typing errors can also help date a typewritten document. Wite-Out Company first introduced this product to the market in 1965. In 1984, Liquid Paper introduced colored correcting fluid to mask corrections on different colored paper stock. The presence of these materials on a typewritten document before their respective introductory dates will strongly suggest a document has been backdated.

Correcting fluids are complex substances composed of different resins, plasticizers, pigments, solvents and binders. The manufacturer of a correcting fluid can be identified by extracting a sample from the document; analyzing it by infrared spectroscopy and comparing the result to a database of correcting fluid spectra. Once known, the manufacturer can be contacted to determine when a particular correcting fluid formulation was first produced. Of course, a correcting fluid could not have been applied to a questioned document before its date of introduction.

Photocopiers

Photocopied documents that suddenly surface during a litigation are often regarded with suspicion. In some cases, these documents are genuine but in other instances, they are produced at the last moment with an astonishing story that they were just discovered recently by some strange coincidence. The subject of interest in these cases is not when the original document was produced but rather the date or period it was photocopied. Three facets of photocopied documents that have dating significance include: the copier technology used, the presence of copier defects, and the properties of the toner and/or paper.

Copier technologies

Just as milestone events in the development of the typewriter are useful for dating purposes, the date of a copied document can be checked against the release date of a particular office technology to insure it was available when the document was allegedly produced.

Different copier technologies include: (1) dual spectrum; (2) stabilization; (3) diffusion transfer; (4) indirect electrostatic; (5) diazo; (6) dye transfer; (7) direct electrostatic; (8) thermographic; and (9) laser. A questioned copied document should be checked to insure its date follows the introductory date of the technology used to produce it, keeping in mind that introductory dates may vary from region to region.

Examination of defects

The most straightforward means of dating photocopied documents relies on defects, 'trash marks' or small flecks of toner that appear in 'white' areas of a copied document. These marks can originate from dirt, foreign material or defects on the glass, platen cover or photosensitive drum of the photocopier (**Fig. 5**). Scratches to the glass or drum, tend to be more permanent and will generate marks on copies produced by a machine until such time as the defective component is removed and replaced. The temporary nature of other defects, such as those originating from dirt or foreign material on the glass, lid or internal components, are temporary in that they can be removed by cleaning the copier surfaces. Genuine photocopied documents made by the same copier that produced the questioned document provide an excellent means of confirming its date. Logs and service records maintained by repair technicians are also helpful in that they often contain photocopies produced before and after copier repairs were made.

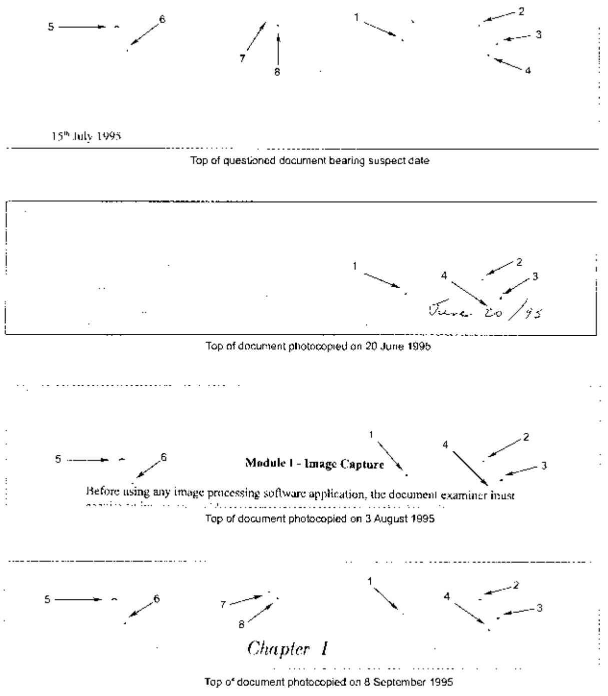


Figure 5 The combination of photocopier 'trash' marks on the questioned document (top) emerged during the period August 3, 1995 to September 8, 1995 and could not have occurred on July 15th, 1995 when copies of the questioned document were allegedly prepared.

Toner analysis

Most photocopier toners consist of: a pigment (usually carbon black); a binder which fixes the pigment to the paper (usually an organic resin such as polystyrene); and additives used to improve the properties of the toner. When any of these components are changed, the event can provide a useful means of dating photocopied documents. Analysis of photocopier toners by infrared spectroscopy and scanning electron microscope equipment with energy dispersive spectrometry can yield information about the chemical and physical properties of toner. A comprehensive library of toners can be used to establish initial production dates. In some cases, the manufacturer will confirm that a particular ingredient was first used several years after the date the photocopy was supposed to be prepared. This would constitute conclusive evidence that the alleged date of the photocopy was false.

The process used to fuse toner to the paper can vary from one photocopier to another. Older photocopiers use cold pressure fusing wherein toner is pressed into the paper surface. Newer generations use either heat alone or both heat and pressure to fuse toner to the surface of the paper. The date a given fusing process first appeared is the earliest that a photocopy bearing this technology could have been produced.

In 1992, it was reported that indentations are imparted to the surface of toner by damage to the surface of a copier's fusing rollers. Fusing roller defects occur through normal wear and tear. They vary with time and consequently the indentations they produce in the surface of toner can be used to estimate when a given photocopied document was produced.

Handwriting and Signatures

The writing of many individuals does not change significantly for most of their adult life. However, despite the constant and repetitive nature of developed handwriting, practically everyone has noticed that their signatures and handwriting do change – especially over long periods of time. The development, progression, and eventual disappearance of handwriting features can be very helpful in solving dating problems. Access to a quantity of specimen material produced during a period of time can show that writers change the shape of certain letters or the form of their signatures (**Fig. 6**). The quantity of

Figure 6 Six signatures produced by a writer during a ten year period show some features that have a temporal significance.

specimens required for this purpose will depend on many factors including: (1) how rapidly the writing changes; (2) what factor(s) influenced the changes; and (3) the number of specimen writings prepared near the period in question. Once the specimens are arranged in chronological order, it is often possible to date a disputed writing within a particular time period.

Rapid changes in a person's writing can result from the sudden onset of a serious illness, the administration of therapeutic drugs or the consequence of a debilitating accident. Although such sudden transitions can create problems for the document examiner, they also provide a means of determining when a questioned signature or handwriting might have been produced.

Contents of a Document

Proof that a document was backdated or postdated can occasionally be found within its contents. These details are often overlooked by the perpetrator as his attention is focused on producing a document that contains the right information. Names, addresses, postal codes, phone numbers, trade names, and job titles mentioned in a document might provide evidence that it was produced at a different time.

Events are occasionally mentioned in correspondence that did not occur until months or years after the date appearing on the document. Verb tenses in relation to events mentioned can also indicate a document was prepared after its purported date. When preparing a postdated or backdated document, the writer may not remember what verb tense to use. Such inconsistencies, especially when repeated, provide a good indication that something is amiss.

When preparing business correspondence, the typist's initials are often placed at the bottom of the document. In fraudulent documents, the initials of a typist who is currently employed by a company may be used instead of the person who held the position on the date that appears on the document.

Computer-printed Documents

Dating computer printed documents is approached in much the same manner as dating typewritten documents. The debut of computer printer technologies are all associated with a date of introduction. Consequently, any document produced by a daisy-wheel, dot-matrix, inkjet or laser printer cannot bear a date that precedes the respective periods when these printers first appeared on the market.

Daisy-wheel printers

The daisy-wheel printer, using a similar impact technology to the typewriter, bridged the gap between typewriters and later generations of computer printers. Although very popular during the 1970s, few daisy-wheel printers are still in use today. The print elements of these machines contain a full set of characters positioned on the end of long spokes attached to a central hub. As the elements spin on a central shaft, the characters are struck at the appropriate time from behind with a plunger. The action of the character striking the paper through an inked ribbon produces a letter on a document.

Like their typewritten counterparts, documents produced by daisy-wheel printers can be dated by considering irregularities in the alignment of letters or damage to their outlines through wear and tear. The source of other temporal defects can be traced to faulty moving components of the printer. These changes provide a means for dating the work of a particular printer. It should be kept in mind, however, that daisy-wheels can be easily removed, discarded and replaced by a new element. All defects associated with the old daisy-wheel will disappear and only those that relate to the printer will remain.

Dot-matrix printers

Dot-matrix printers gained popularity during the early 1980s. Early models had nine metal pins arranged along a vertical axis that struck the paper through an inked ribbon while the printhead moved

across the page. At the end of a single pass, the paper would advance slightly and the printhead would return across the page in the opposite direction. This process would be repeated until the entire page was printed. Printing produced by dot-matrix printers improved as 12, 18 and 24 pin models became available. These produced sharper printing which was referred to as 'near letter quality' or NLQ printing. The dates when these progressive improvements occurred provide a further means of limiting computer printed document to a particular period.

Documents printed by dot-matrix printers can also be dated by the sudden appearance of printing defects which are due to broken or bent pins, worn printhead housings, or other manifestations caused by defective printer components.

Ink-jet and laser printers

Documents produced by inkjet or laser printers could only be produced after these technologies were introduced. A computer-generated document can often be associated to a particular printer manufacturer based on the presence of class characteristics. The chemical composition of ink-jet ink or toner can also be useful for determining if a document has been backdated.

All nonimpact computer printers use computer software to generate printed characters. Printer control language (PCL) defines how letters belonging to a particular typestyle are shaped. For example, until October 1993 no Hewlett Packard printer control language was capable of handling 600 dots per inch (dpi) printing The Hewlett Packard LaserJet 4, introduced in October 1993, was distributed with a special internal Courier font developed specifically for 600 dpi printing. This typestyle was different from any prior Courier font used in LaserJet printers (**Fig. 7**). Since the LaserJet 4 was introduced in October 1993, any document which contains this special Courier font but dated earlier than this must be backdated.

The mechanics of laser printers are very similar to the processes used by modern photocopier machines. Hence, methods for dating photocopied documents described above also apply to documents produced by laser printers.

Facsimile Documents

Recently, facsimile machines have become a common form of business communication. Although the first fax was designed by Alexander Bain and patented during 1843, the machine has only really gained popularity since the early 1980s. Facsimile documents are often presented as proof that business transactions or agreements took place on a particular date. Not all

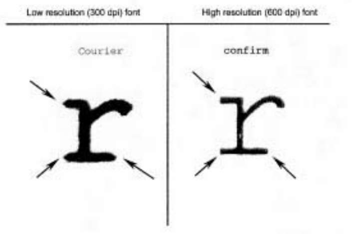


Figure 7 Letters 'r' produced by a HP LaserJet III printer (left) and by a HP LaserJet 4 introduced to the market in October 1993 (right) show several conspicuous differences between the internal Courier fonts installed on these printers

of these documents, however, are genuine. Fast and convenient for their users, facsimile machines also provide fraud artists with an opportunity to fabricate documents and defraud unsuspecting victims.

The Transmitting Terminal Identifier (TTI) header usually appears at the top of most facsimile documents. This header may contain the page number, the date, the time the message was sent and other information supplied by the sending machine. Although dispatched by the sending machine, this information is printed by the fax that receives the message. A Receiving Terminal Identifier (RTI) printed by the receiving fax machine can also appear at the bottom of transmitted documents. The TTI and RTI of every suspected faxed document warrant close inspection.

In many cases, the date and time appearing in the TTI of faxed message are correct. It should be noted, however, that these settings can be quickly changed by anyone who has access to the machine and who possesses the knowledge to make the adjustments.

It is possible to identify the make and model of both sending and receiving machines by comparing the TTI and RTI of a received facsimile document to a library or collection of fax fonts. If such a search indicates that one or both facsimile machines were not available when the questioned fax was received, then it casts suspicion on the authenticity of the transmitted document. If long-distance charges were incurred when sending the facsimile transmissions, telephone records should be examined for evidence that a fax was sent on the date and time alleged. Telephone numbers and area codes appearing in the TTI or on the cover sheet should also be checked to insure they existed when the document was supposed to have been sent.

The format and content of the fax cover sheet should be examined to determine if it is consistent with that used during the period in question. Details of the transmitted message and cover sheet should also be examined to insure people's names, titles or initials are appropriate for the period in question.

Other factors worthy of consideration can be found under the 'Contents of a Document' heading of this chapter.

Cachet Impressions

The development of rubber stamps followed the discovery of vulcanizing rubber by Charles Goodyear. The first commercial production of rubber stamps occurred in 1864. Since that time, the processes used to manufacture stamps have undergone several improvements as the demand for better quality rubber stamps increased. The first pre-inked stamp, Perma Stamp was produced in 1958. These stamps are still a popular item in stationery stores. Although today's stamps are still referred to as 'rubber stamps', most stamps are now produced from a plastic-based photopolymer material.

Both rubber and plastic deteriorate over time. The relief edges of a stamp can crack or break off, an ink/dirt mixture can clog deep crevices and the relief areas of a stamp can become worn through constant use. These events introduce flaws that are reproduced in the impressions produced by a worn stamp. The approximate period when a stamp impression was made can be determined by comparing its defects with standards from the same stamp arranged in chronological order.

Another method by which stamp impressions can be dated involves changes to the size of some stamps with time. It has been found that stamps can shrink as much as 1.5 mm during a four-year period. Although this phenomenon is relatively rare, it does provide yet another means of dating stamp impressions.

Glues, Tapes and Paper Fasteners

Adhesives used to manufacture envelopes, stationery pads and tapes occasionally undergo changes or modifications to improve their properties. Such changes can be used to establish the earliest date that a document manufactured with a given adhesive was produced. The stationery manufacturer or adhesive company should always be contacted to verify the date when a particular adhesive was first used.

Lift-off tape was introduced by IBM to facilitate the correction of typewriting errors. This innovation, first introduced to the market by IBM on the 1st of April 1973, removed unwanted typed characters by overstriking letters through the lift-off tape. This action would lift the letter from the document and allow the typist to correct errors with little disturbance to the paper surface.

Indented Writing

Indented handwritten impressions made in the surface of a document can reveal important information about whether written entries on a piece of paper were made before or after the indented writing occurred. Such sequence determinations are confirmed by subjecting the document to an ElectroStatic Detection Apparatus (ESDA) examination.

It is often possible to establish the exact date when indented handwritten impressions on a document were produced. An ESDA examination that establishes the visible writing on a questioned document was made after dated indented impressions can provide an unusual but effective method for confirming the document was backdated.

Handwritten entries in a journal, ledger, note pad or receipt book usually produce indented impressions on underlying sheets of paper. If it is necessary to date one of the sheets which was removed, its original location can be confirmed by matching writing on the document with corresponding impressions on the other bound papers. If the dates on adjacent pages are reliable, this simple method enables the document examiner to place the questioned document within a particular time frame.

Guillotine Marks

The exposed edges of receipt books, reams of paper and stationery pads may contain marks produced by cutters or guillotine blades used to trim these products to size. These stria, often referred to as 'guillotine marks', do not run perpendicular to the surface of the paper but run at an angel across the trimmed surfaces. Their locations along the four edges of a document can indicate where a sheet was positioned in the original stack of paper.

Access to several documents from the same stack of paper is needed to establish a cutting pattern against which the contested document will be compared. Once the location of guillotine marks on the four edges of the questioned sheet match the position of sheets from the same lot, any dating information on adjacent sheets can be used to determine when the questioned document was written. If the questioned document is not contemporaneous with information on adjacent sheets of stationery, some plausible explanation should be sought.

Summary

Many methods can be employed to determine if a questioned document's date has been falsified. People who fraudulently alter or misrepresent the date of a

document are usually careful to ensure its general appearance will meet close scrutiny. Consequently, it is better to approach document dating problems from a broader perspective rather than focusing attention on those areas which immediately arouse suspicion. Any decision to limit the scope of forensic investigations in the interest of saving time or money should be weighed carefully. Solutions to document dating problems are often dependent on an inspection of all areas of a document for details confirming whether it was prepared on a given date. Such an approach will insure that a thorough investigation is carried out and crucial evidence will not be overlooked.

See also: **Document Analysis**: Handwriting; Analytical Methods; Forgery/Counterfeits; Ink Analysis.

Further Reading

Blackledge RD and Gernandt MN (1993) The pH pen – a means of comparing paper products. *Journal of Forensic Sciences* **38**:134–142.

Brunelle RL (1995) A sequential multiple approach to determining the relative age of writing inks. *International Journal of Forensic Document Examiners* **1**:94–98.

Brunelle RL (1992) Ink dating – the state of the art. *Journal of Forensic Sciences* **37**:113–124.

Brunelle RL and Cantu AA (1987) A critical evaluation of current ink dating technique. *Journal of Forensic Sciences* **32**:1522–1536.

Brunelle RL and Reed RW (1984) *Forensic Examination of Ink and Paper*. Springfield: Charles C. Thomas.

Cantu AA and Prough RS (1987) On the relative aging of ink – the solvent extraction technique. *Journal of Forensic Sciences* **32**:1151–1174.

Cantu AA (1996) A sketch of analytical method for document dating Part II. The dynamic approach: determining age dependent analytical profiles. *International Journal of Questioned Document Examiners*. **2**:192–208. (Errata in **2**:570–571)

Gerhart J (1992) Identification of photocopiers from fusing roller defects. *Journal of Forensic Sciences* **37**:130–139.

Godown L (1969) Forgeries over genuine signatures. *Journal of Forensic Sciences* **14**:463–468.

Kelly JH (1983) *Classification and identification of modern office copiers*. Houston: American Board of Forensic Document Examiners Inc.

Nickell J (1990) *Pen, Ink, & Evidence*. Lexington: The University Press of Kentucky.

Osborn AS (1929) *Questioned Documents*, 2nd edn. Albany: Boyd Printing Co.

Purtell DJ (1980) Dating a signature. *Forensic Science International* **15**:243–248.

Starrs J (1991) The case of the doctor who doctored the documents. *Scientific Sleuthing Review* **15**:1.

Totty RN (1990) The examination of photocopy documents. *Forensic Science International* **46**:121–126.

Forgery/Counterfeits

P W Pfefferli, Kantonspolizei, Zurich, Switzerland

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0478

Introduction

Forensic document examination is a complex matter requiring various areas of expertise to identify document forgeries of all kinds. Although in the modern office the electronically mailed document is of growing importance, the traditional (paper) document still has its place, and thousands of different kinds are used for documentation purposes. The threat of illegal document forgery exists and will always exist.

The examination of questioned documents is an ongoing challenge to forensic sciences, to cope with the modern technology of document manufacturing and the increasingly sophisticated document forgeries. The many facets of modern documents demand many areas of expertise from forensic specialists in order to identify and authenticate a questioned document. The identification of handwriting is a discipline of its own, requiring for appropriate trained experts to compare and identify handwritings. The aim of the examination of technical typewriting and printing produced by office machines is to determine typestyle and origin. In addition to the written or printed content, it is necessary to examine other document components. Ink and paper examination, by nondestructive and analytical methods and techniques, help to determine its authenticity with respect to origin. These areas are equally relevant in general questioned document examination, but when documents of high value and protected by the law are fabricated for criminal purposes, the forensic document analysis may be of even higher importance and special expertise is required. Counterfeiting and forgery of value and security documents occurs on a large scale.

The Threat

Counterfeited documents are documents that are reproduced as originals. The entire document is a fake, e.g. counterfeited banknotes, bonds, checks, vouchers, driving licenses, lottery and admission tickets, stamps, certificates etc. If the reproduction is made by using materials and procedures which do not correspond to the original, the counterfeit is called an imitation.

Forged documents are altered originals produced by adding, removing or substituting relevant information or features, for example a modified check

amount, an imitated signature on a contract, altered passport or gift certificate, altered payment card or credit card.

The counterfeiting of banknotes, bonds, checks or driving licenses and the illegal reproduction of passports and travel documents and check and credit card fraud are high profile crimes, committed mostly not by an individual, but by criminal organizations involved in terrorism, illegal cross-border traffic, drug trafficking, financial transactions or other activities of organized crime. By definition, such highly valuable so-called security documents, such as banknotes and passports, should be specially protected against misuse. The higher the value of a document (absolute or relative) and the greater the damage caused by fraudulent actions, the higher the risk of fraud. Therefore, there will be greater protection of the document against fraud. However, a great number of these documents are still too easy to counterfeit, because they do not meet the minimum standard of manufacturing. If a banknote of reasonable quality can be reproduced by the use of modern desk top publishing or simply with a full color copier, then document securities are missing. If a security document, such as a passport, can easily be altered by substitution of the passport owner's photograph or by chemical eradication of the passport entries, the document securities of the passport do not meet the necessary standard to prevent counterfeiting and forgery. The threat of document counterfeiting and forgery has become twofold, due to the quantity and the quality of counterfeited and forged documents.

The increase of fake documents of all kinds, the flood of low-profile security document fraud, mainly for immigration and cross-border purposes as well as the frighteningly good quality of some seized currency counterfeits, give rise to concern, which is much greater than just a forensic dimension.

Document fraud by counterfeiting and forgery is no longer a form of crime exclusively encountered in white collar crimes or secret intelligence activities. Document counterfeit and forgery is nowadays found in all forms of crime. Statistics regarding the most often counterfeited currency worldwide – the US dollar – or the number of seized fake travel documents at border controls and immigration services, show frighteningly high figures for these trans-national phenomena. The relevant question of the document examiner 'genuine or false?' is certainly not trivial. What is considered genuine and what is considered fake? Even valuable documents are too often of such poor quality that it is not particularly difficult to imitate them.

The improved technology of modern office machines has meant that almost perfect reproductions of an original can be produced. This is unfortunately true not only for low profile documents used in daily business, but also for security documents. The full color copier has become the most often used tool for illegally reproducing value documents. Used as a printer device linked with a computer, it offers a wide range of technical features to match the copy to the original. Colors can be balanced, numberings changed or signatures scanned. Even some of the traditional security features, for example the security thread of banknotes, can easily be imitated.

Banknotes are documents which are normally not altered, but counterfeited, however, the situation is different in the case of check fraud. Bank checks and personal checks are value documents which are, or should be, protected against misuse as banknotes are. However, the level of securities is mostly considerably lower, making it even easier for criminals to forge. Besides reproducing the entire check from scratch, the threat of check fraud comes from altered written checks and forged stolen blank checks. The main obstacle to altering an already written check is the erasing and substituting of entries without revealing the changes. In the case of forged blank checks the important challenge for the forger is to imitate the check owner's signature.

Another category of faked documents, similar in complexity to check fraud, are counterfeited and forged identity documents. The attraction to the criminal forger is – unlike banknotes and checks – not the pecuniary value, but the ability to create a new identity. Since modern identity documents, such as passports or ID cards and driving licenses, are complex documents with respect to manufacturing and technical set-up, it is easier and quicker to alter genuine documents than to counterfeit an entire document. Statistics from immigration services forgery desk examination make it clear that among the hundreds of different ID documents used worldwide, there are still too many in which the substitution of the identity of the document holder is not a difficult task for a skilled forger. As for check fraud, one of the forger's difficulties is the alteration of typewritten and/or handwritten entries. Moreover, the document holder's photograph must be substituted, quite often this is not a difficult task, since the weakest link of most easily altered ID documents is the document's protection against substitution of the photograph. Many of these documents have ring staples and dry or wet stamp cachets to secure the photograph, which do not give any protection against substitution. Even thin-laminate sealed photographs are no problem for the professional forger who has little difficulty in splitting the photograph substrate in half to get access to the picture without breaking the laminate.

A final category of document complexity are the machine-readable documents, such as payment cards and credit cards, and also plastic card legitimization documents, for example modern identity cards, driving licenses and the various types of smart cards. The threat of misuse of plastic money and plastic cards is inescapable and this form of fraud is increasing dramatically. Small-scale forgeries involve the altering of stolen cards. In addition to altering the entries on the plastic substrate, the manipulation of the stored information on the magnetic strip, the electronic chip, is an additional difficulty. However, this obstacle can be overcome if necessary. The threat of today's card fraud is, however, the amount of organized crime involved in forgery and the counterfeit of plastic cards. The criminal organizations produce blank cards with machine-readable devices, which are then delivered to the customer for encoding and, if necessary, to be completed by printing, embossing or additional optical security devices (for example hologram).

It is true that despite the high quality of some professionally produced forgeries and counterfeits of security documents, the forensic specialists have little difficulty detecting the fraud when checking a questioned document thoroughly. The recipient of these counterfeited documents, however, is likely to be an unskilled and not technically equipped customer.

Document security

It goes without saying, that there are different classes of documents, according to the probability of fraud, the damage caused by it and the absolute as well as the relative value of the document. The class which is best protected against counterfeiting and forgery is high-value documents, such as banknotes, passports, checks and credit cards. Whereas other value documents, such as stamps, vouchers, certificates, admission tickets, stamps, lottery tickets, visa documents, permits and many others may be considered as a class of lower security value and therefore also of lower document security protection. The understanding of the counterfeiting and forgery potential of different types of document demands an enhanced know-how of the state of the art of technical document securities and the respective possibilities and limits. Generally speaking, there are three approaches to improve the technical security of a document against counterfeiting and forgery.

First there is the use of exclusive materials and controlled substances: the higher the value of a document, the more exclusive the materials for manufacturing should be. Controlled substances should be produced by a restricted number of manufacturers and not normally used for the production of common documents; for example paper from specialized paper mills, paper containing special fibers, papers without optical brightener, multilayer polymers for synthetic documents, printing inks of particular spectral composition, printing inks specially resistant to chemical eradication or alternatively, especially sensitive to chemical attacks etc.

Secondly security printing techniques should be used. Besides offset as the most common technique used in printing, special printing techniques such as letterpress, intaglio (relief printing) and silk-screen are highly effective and not easy to imitate.

Thirdly, additional security attributes should be used. There are various security devices to improve the protection of a document against counterfeiting and forgery. These are used either in combination with the document materials during manufacturing or added on the printed document. These modern security devices include (among others):

- Paper securities: watermarks; fluorescent fibers and planchettes;
- Printing securities: guilloche pattern (printed background pattern of interlacing fine lines); simultaneous printing (perfect fitting of recto/verso printing); fluorescent and magnetic inks; optically variable inks (ink taking on different colors depending on the angle of viewing); rainbow printing (continuous change of colors); microprinting (extra small printing to be seen only after magnification); latent images (visualization depending on angle of viewing); scrambled indicia (scrambled latent image, only visible through a special lens);
- Securities of personalized entries: text printed by use of matrix printers or laser printer; digitized laser printing of text, photos or signatures;
- securities to prevent substitution of pages, photos: staples; embossing stamps (dry stamps), stamp cachets; optically variable devices OVD (holograms, kinegrams, latent images); protective films (thin laminates) with ink printing, retro-reflective-pattern, diffraction grating image; perforations;
- machine readable securities: OCR printing; magnetic strips.

Methods of Examination

The methodology of document counterfeit and forgery analysis is first a question of where the authenticity checking takes place. Depending on the form of fraud and the type of document involved, a first examination will not necessarily be done in the document laboratory, but at the first line of

inspection; for example bank institutions, immigration services, border controls, social security services etc. Therefore, the personnel should be capable of identifying a false document (in the time available and technical support) before sending it to the forensic document laboratory for thorough checking. For suspect documents, that are not identified by a first line of checking, the entire examination process will be handled by forensic experts to confirm or refute the suspicion.

There are three levels of detection of counterfeited or forged documents, depending on the authenticity features to be checked.

Level 1 includes features, which can be visually examined without special equipment, such as watermarks, security threads, relief structures, latent images, stamps (dry stamps or stamp cachets), OVD (e.g. kinegrams), mechanical erasure, perforations of staples, residues of adhesives, obliterated writing and other visible traces of manipulation.

Level 2 refers to features to be checked with technical facilities such as: visible light to look at printing quality, traces of substitution or relief effects; UV light to detect UV-fluorescent securities (paper bleaching, fibers, printing) and UV-visible forgery traces (chemical eradication); infrared radiation to examine writing inks; reading devices to detect magnetic printing; special viewing systems for retroreflective securities; reading systems for machine readable devices.

Level 3 examinations have to be carried out in the document laboratory, with sophisticated equipment that can not be used for field examination.

- Electrostatic examination of latent impressions of handwriting and detection of indentations (indentations of erased writings, traced signatures on checks or passports);
- Juxtaposition as well as superposition comparison with the microscope or by digital image processing, to visualize faint printing defects;
- High magnification microscopy examination to detect trace evidence of mechanical erasure or substitution (residues of writing and printing ink, adhesives; traces of cutting, evidence for manipulation of staples, binding, stitching, etc.);
- Infrared absorption and infrared luminescence examination to decipher nondestructively faint or obliterated writing and printing (traced signatures, modified stamp cachets);
- Microspectrophotometric (colorimetric) analysis of the spectral composition of colors;
- Chemical detection of chemical erasures;
- Chemical spot reactions to check the organic and inorganic composition of document materials (paper, ink);
- Spectroscopic examination of organic compounds (polymers of synthetic documents, e.g. credit cards; identification of photocopy toners).

The methodology follows the general system of forensic examination and standard procedures for document-related examination, based mainly on the detection of irregularities or by comparing the questioned item with an authentic specimen. An absolute opinion on the authenticity of a questioned document may, however, not always be given, even after high technology examination, simply because the differences between the original and the suspected imitation are not conclusive. For example, the authentic document may be of poor (print) quality, or the highly dangerous forms of fraud and forgery are not known.

Comprehensive reference manuals and databases for highly protected security documents, particularly banknotes, passports and visas, try to keep updated collections of images from authentic documents, including close-up views and specifications of relevant security features. In addition document experts of immigration and forensic science services are networking information on detected counterfeits and forgeries. This is a valuable help for fast authenticity checking, mainly at the front line of inspection. For many other documents however, this type of reference material does not exist. This is also the case with documents that had been produced in such a variety of 'official' types, that it is no longer clear what should be considered authentic!

Another factor that is making the unambiguous identification of suspected questioned documents more and more difficult, is the improvement in office technology; these are also available for use in forgery and other criminal activities. Digital scanners and printers have reached a technology level able to produce documents, even security documents, of a quality standard, which makes it more and more difficult to differentiate between genuine and fake. This is particularly true for laser color copy technology with its almost unlimited software possibilities. However, combating color copy fraud is possible and technologies to prevent it are available. Either by programming the scanning software, to recognize automatically unauthorized reproduction, e.g. of banknotes and checks, or by printing on the copy a latent machine-readable code, which can lead via the manufacturer to the customer's copy machine.

The detection of document counterfeits and forgery is a challenge not only to forensic sciences!

See also: **Forgery and Fraud**: Counterfeit Currency; Payment Cards. **Document Analysis**: Handwriting; Ink Analysis.

Further Reading

ASTM Standard E 1422-98 *Standard Guide for Test Methods for Forensic Writing Ink Comparison*. West Conshohocken: The American Society for Testing and Materials (ASTM).

Ellen D (1997) *The Scientific Examination of Documents*. London: Taylor & Francis.

Interpol (ed.) *Counterfeits & Forgeries*. Amsterdam: Editions Keesing Publishers.

Passport Handbook. Amsterdam: Editions Keesing Publishers.

Rapp B (1990) *Counterfeit I.D. Made Easy*. Port Townsend, Washington: Loompanics Unlimited.

Rapp B (1991) *Credit Card Fraud*. Port Townsend, Washington: Loompanics Unlimited.

Rapp B (1991) *Check Fraud Investigation*. Port Townsend, Washington: Loompanics Unlimited.

Van den Assem B, Brongers D, Rath J *et al.* (1994) *Security Documents – Practical Guide for the Security Printing Sector*. The Hague: Sdu Publisher.

Van Renesse R (ed) (1998) *Optical Document Security*. Boston, London: Artech House.

Handwriting

M Vos, S Strach and **P Westwood**, Forensic Document Services, PO Box 543, Queanbeyan, NSW, Australia

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0475

Introduction

Handwriting can be described as the formation of letters, characters or symbols, using a writing implement, according to a recognizable pattern which is designed to communicate with another person. The issue of who wrote a particular piece of handwriting can be central to many legal proceedings. Accepted ways to prove handwriting include admission by the writer or the testimony of a person who witnessed the writing. If the writer is uncooperative or not available, and there are no apparent witnesses, or the writing is otherwise in dispute, then there is a need to resort to other means of identification. There are two other means acceptable by the courts. The first is to call a person acquainted with the writer's handwriting; the second, and more reliable, is to employ the services of a forensic document examiner to undertake a comparative examination of the writing in question with known specimen handwriting of the suspected writer of the questioned material. A document examiner may be asked to examine such diverse documents as anonymous letters, writings and/or signatures on cheques, credit cards and vouchers, wills, mortgages, other legal documents, medical records and diaries. It should be noted that, although documents generally consist of writing on paper, the forensic document examiner can be called upon to examine writings on other less conventional documents such as blackboards, whiteboards, pieces of pasta, body parts, fence posts and graffiti on buildings, walls, windows etc.

Writing is taught from a model system. The act of continuously repeating a written character fixes the form of that character in the mind of the writer, normally during childhood, until the production of this form becomes 'automatic'. From the moment people start learning to write, they introduce deviations from the model writing system taught. The extent of these deviations increases as the writing style becomes more personalized, resulting in a style which is the product of many factors including the model system, artistic ability, muscular control, nature of employment, frequency of writing and exposure to the writings of others. This results in an individual writing style, the development of which occurs throughout the childhood and adolescent years, and often beyond (**Fig. 1**).

A similar evolution can be seen in the development of a person's signature with the individual practicing a certain style which changes over time under the influence of the factors described above, until an often quite individualized pictorial representation of the person's name is formed.

Document examiners refer to writing characteristics attributable to a model system as being 'class' or 'style' characteristics. Different styles of writing are taught in different countries and sometimes regions, which accounts for some of the variability in handwriting styles. The styles taught also change over time. For example, in Australia today it is unusual to be taught the elaborate cursive style of writing which was once commonly taught. A number of terms are used by document examiners for characters found infrequently in the general population including 'unusual', 'personal' or possibly 'individual' characteristics. The term 'individual' characteristic is somewhat misleading if applied to one letter form, since one letter form even if very unusual is not sufficient to individualize, or identify, the writer. In order to distinguish the writing of any one person, the handwriting expert builds a background knowledge over time, from examination of the writings of many different people, of what can be considered common

Figure 1 Ten people's writing of the place name 'Queanbeyan', illustrating the differences and occasional similarities which arise in the writings of different people.

or unusual features of handwriting and how handwriting can change under different circumstances. If sufficient handwriting is present, it is the combination of rare and common features which serves to identify a particular writer.

Although we use the terms forensic document examiner and handwriting expert interchangeably for the purposes of this article, it should be noted that the term 'handwriting expert' is sometimes used by those who seek to determine a person's psychological characteristics from their handwriting (graphologists). Caution should be taken when engaging the services of a 'handwriting expert' to ensure that the person engaged is fully trained and qualified in the field of forensic document examination. Although the bulk of the work of most forensic document examiners focuses on the scientific comparative examination of handwriting and signatures, there are document examiners who almost exclusively examine aspects of documents other than handwriting.

Handwriting Comparison

When a particular piece of handwriting is called into question, the handwriting expert will examine the writing in question visually, with the aid of hand held magnifiers and with a microscope. Features such as the details of construction of individual letters and letter combinations, comparative height and size relationships of particular letters and letter combinations, alignment of letters and words, slope, speed of writing, angularity, pressure, shading, diacritics and layout are considered.

Constructional details such as line direction and sequences of separate strokes can often be determined from microscopic examination. Evidence of the sequence of separate strokes can sometimes be obtained by the direction of any small terminating and commencing strokes between separate parts of a character, called spurs or ticks, or faint connections between these parts, showing the direction of pen travel just before a pen lift, or just after the pen has contacted the paper following a pen lift. Usually, examination of several examples of each letter are required to make such a determination. Ball-point pen striations can also be used to determine line directions as such striations run from the inside to the outside of a curved stroke (**Fig. 2**). The striation pattern present at the conclusion of a stroke, may be retained on the ball of the pen and transferred to the beginning of the next stroke. This 'memory effect' can be used to determine the sequence of strokes. Other characteristics of the ball-point pen stroke that allow determination of direction of stroke include 'gooping' (an extra heavy deposit of ink) following a curve and dried or reduced ink starts at the beginning of lines. For this reason, it is strongly recommended that at least some of any requested set of specimen handwriting is written with a ball-point pen. Higher power microscopic ($\times 30$ or more) examination can also

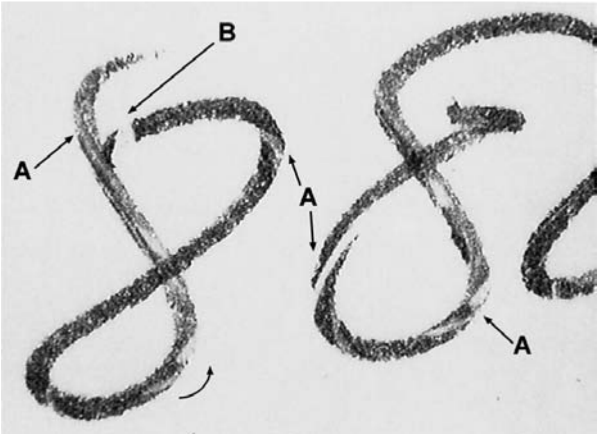


Figure 2 Ball-point pen writings showing (**A**) striations and (**B**) reduced ink start. Arrows indicate direction of pen movement.

sometimes be used to determine line direction of ball-point pen writing, and almost always for pencil writing. This can be determined from the build up of ink or graphite predominantly on the 'upstream' side of the paper fibers which faced the approaching writing tip.

The known writing of the person suspected of writing the questioned document is examined in a manner similar to that employed with respect to the questioned document. Specimen handwriting needs to be comparable in style with the questioned handwriting. For example, questioned upper case block printing should be compared with specimen upper case block printing. It is also preferable to examine specimen writing on forms of a similar printed layout. Once suitable specimen handwriting is obtained, an initial comparison is made between the known writing samples to ascertain if they can reasonably be taken to have been written by the one person; it can happen that documents which are said to bear the known writing of one person may include (at times unwittingly) the writing of other persons. This is a particular consideration when using diaries, address books or other documents where many persons may have had access to and written on the document. From the examination of specimen writing, the expert assesses how the writing of this particular individual varies within itself.

Once the requirements for suitable specimen handwriting are met, the questioned and specimen handwritings are compared with each other, and an assessment made of similarities and differences between the two writings. A character in the questioned writing is considered to be similar if it, or its range of variation, falls within the range of variation for this character in the specimen handwriting. If the character or its range of variation in the questioned writing falls outside the range of variation seen in the specimen handwriting (for example there are differences in form (shape) or in direction, number or sequence of strokes), then this is considered to be a difference. Some differences are regarded as more significant than others; for example, the repeated appearance of a different sequence or direction of strokes for a naturally written block capital letter (**Fig. 3**) or numeral may be regarded as a more fundamental difference than a shape difference of a lower-case cursive letter. If significant differences are found, this usually results in a conclusion that the two writings are unlikely to have been written by the same person or, alternatively, there is nothing to link the questioned and specimen writings as being by one person. Although definite negative conclusions are sometimes justified, they should be used sparingly and in special circumstances. This is because of the possibility which should be considered, although remote, of a person having two distinct handwriting styles. With similarities also, some are more significant than others. The significance is greater if it involves a handwriting feature found more rarely in the appropriate general population. This is usually assessed subjectively, based on the experience of the examiner in observing handwriting features in very many writers. The keeping of handwriting reference collections, and using such references to determine the rarity or otherwise of a particular feature, is becoming more prevalent and can provide some statistical backing for the results of handwriting comparisons.

Finally the examiner, before arriving at a finding, assesses the results of the comparison process in terms of similarities and/or differences in all comparable features of the handwriting and their significance. If no significant differences are found, the examiner assesses the likelihood that all of the similarities could have occurred by chance in the writings of two people, or alternatively could have originated in one person simulating the handwriting of another without leaving any evidence (see below) of the simulation process. These are usually assessed as subjective probabilities, based on the premise (from the probability multiplication law for independent events) that the probability of very many similar handwriting features, not necessarily individually rare, occurring in combination in the writings of two people is considered extremely small or negligible. Where the probability of a chance match in the writings of two people and the probability that another person has successfully simulated the writing style of the writer of the specimens are both considered negligibly small, the document examiner reaches an unqualified conclusion that the writer of the specimens wrote the

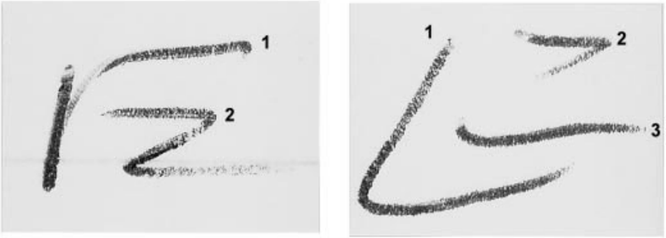


Figure 3 Two forms of the capital letter 'E' showing different sequences of strokes indicated by numbers.

questioned entry. Such a conclusion may on occasion be reached even where some apparent differences are found provided the document examiner is satisfied that these apparent differences are due to disguise, accidental formation or the use of an alternative letter form. Where the writer has made a correction or an accidental stroke caused, for example, by jolting of the pen, the character formed may vary from other examples of that character within the body of the text.

On the other hand the document examiner may reach a conclusion expressed in terms of some degree of probability if doubts arise from:

1. the paucity of questioned or known writings;
2. the inherent quality of the document;
3. the disguised nature of the questioned or known writing;
4. the lack of contemporaneity between the questioned and known writings;
5. the lack of individuality in the writing; or
6. any other reason.

Such qualified findings may nevertheless be of assistance to investigators, lawyers and to courts on many occasions. It should be noted that these are subjective, not mathematically determined, probabilities and actually refer to the document examiner's perceived probability of being correct in pointing to the writer of the specimens as being the writer of the questioned material.

There is no fixed amount of writing required for a definite handwriting conclusion, or other level of certainty. This is because the overall unusualness of the combination of features which makes up a person's handwriting varies from person to person. Together, these character forms and other handwriting features make a person's writing in most cases unique and identifiable. Where a person's writing does not vary much from the model system taught, considerably more questioned and specimen writing is needed in order to reach a conclusion as to whether this person wrote the questioned writing. Conversely, a highly individualistic writing can be identified with less questioned and specimen writing.

Occasionally, because of extreme limiting factors such as severe distortion of the questioned writing, the small amount of writing bearing only common features or inadequate specimens, the document examiner is unable to reach any useful conclusion and expresses the result as inconclusive.

Signature Comparison

In the case of a signature examination the examiner follows similar steps to those used for the handwriting comparison, firstly determining that the specimen signatures provided can be taken to be the signatures of one person, then comparing the specimen signatures with each of the questioned signatures in macroscopic and microscopic detail. It should not be assumed that in the case of multiple questioned signatures in one name that all of the questioned signatures have been written by the one person. Whereas handwriting comparisons usually require a reasonable amount of questioned handwriting, it is often possible to reach a definite conclusion that a highly personalized signature is genuine (**Fig. 4**) with much less questioned material. This is because the likeli-

Figure 4 Varying degrees of personalization in a range of signatures.

hood of a chance 'match' occurring between an idiosyncratic signature and the handwriting of another person writing this signature style without knowledge of the genuine signature is significantly reduced.

For most, but not all signature cases, which involve the comparison of a set of specimen signatures with generally pictorially similar questioned signature(s), the main issue is whether the questioned signature was written by the writer of the specimens, or whether the signature was written as a simulation of the style of the genuine signature. It is not possible to determine by handwriting/signature comparison methods who wrote a completely simulated signature (or handwriting). In such a simulation the normal handwriting features are distorted by the simulation process. In less complete simulations there may be evidence of the writer's natural handwriting characteristics. For most handwriting comparison cases, the main issue for generally similar writings is whether the writer of the specimens wrote the questioned writing, or whether a chance match has occurred in the writing of two people. It should be stressed that this does not mean that handwriting and signature comparisons are fundamentally different, simply that there is often a different emphasis.

General Considerations for Handwriting and Signature Comparisons

For all handwriting and signature comparison cases, three main hypotheses to explain the observations must be considered.

1. The writer of the specimens wrote the questioned material.
2. A person other than the writer of the specimens wrote the questioned material, any similarities to the questioned writing having arisen by chance coincidence.
3. A person other than the writer of the specimens wrote the questioned material, any similarities to the questioned material having arisen because of a simulation process.

Complications which may further arise are considerations of distortion, disguise or self-simulation of the writing in the case of hypothesis (1), combinations of possible effects of (2) and (3), multiple writers etc. The document examiner needs to consider all feasible possibilities which might explain the observations and be aware of the danger of not considering all of these possibilities.

Disguise of Handwriting and Signatures

There are many types of documents on which disguised handwriting appears, from an innocently sent Valentine's Day card to a note demanding money from a bank teller.

Obvious features of a person's handwriting, such as slope and/or size, are often changed as a form of disguise. Some people will deliberately introduce letter forms that are markedly different from their usual letter forms and some will use the unaccustomed hand in their writing. In the case of the latter, the writing produced can display the effects of poor pen control and appear untidy, but the subconsciously produced letter constructions and proportions may remain approximately the same, with the possible exception of direction of some strokes. The major difficulty with most disguised writing is maintaining the disguise over a length of text. Where a basic disguise may succeed over a few words, the disguise is often forgotten over a number of sentences with the writer reverting to a natural handwriting style.

A totally different style of writing may also be used as a disguise. For example, if a person normally uses printed script, cursive writing may be tried as a disguise. In this instance however, there will often be examples of the alternative style available for comparison. There are, nevertheless, some forms of carefully disguised writing for which determination of the writer from handwriting comparison methods may be difficult or impossible.

Disguised signatures are written by those persons intending to later deny the signature they have written. Generally the signature produced is so close to the specimen signatures, except for one or two differences, that the document examiner will identify the signature as being genuine despite the attempted disguise (**Figs 5** and **6**). Self-simulation of a signature or handwriting as a form of disguise can be significantly more difficult or impossible to detect. The writer of a completely simulated signature or piece of writing may be impossible to determine.

Simulation of Handwriting and Signatures

Simulation of handwriting presents a different set of problems for the potential offender. The person intending to copy the writing of another person needs to obtain some specimen writing of the other person. The copyist then has a number of options. The handwriting can be 'drawn' or some form of tracing can be used to produce the simulated writing. When drawing the writing, the copyist must stop frequently to check the construction of letters and words used by the person whose writing is being simulated, considerably reducing the speed of the writing process. Where no examples of a particular letter are available, the copyist must use some other form. As a result, the writing thus produced is generally slowly written; the words

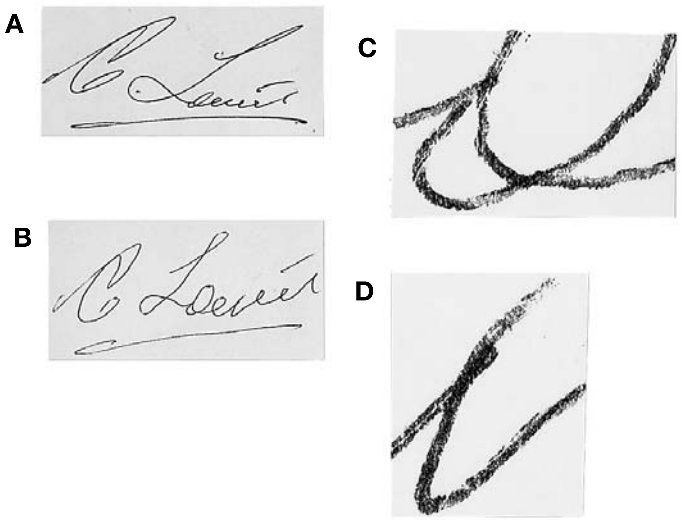


Figure 5 (**A**) Genuine signature 'C. Lewis' and (**B**) slowly written freehand simulation of (**A**); (**C**) detail of the letter 'e' from the simulated signature exhibiting poor line quality; (**D**) detail of the letter 'w' from the simulated signature showing uncharacteristic pen lift.

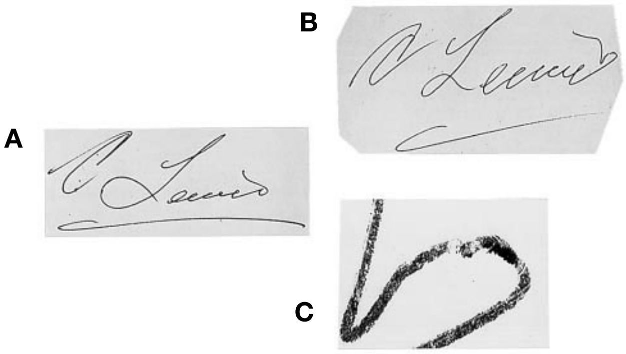


Figure 6 (**A**) Genuine signature 'C. Lewis' and (**B**) quickly written freehand simulation of (**A**). Signature (**B**) shows errors in letter proportion and size relationships and an uncharacteristic pen lift in the final part of the signature (**C**).

may display many pauses in writing where none would usually occur, and the letter forms may display many more variations than the genuine writing since the desire to copy the correct letter conflicts with the copyist's natural way of writing a particular letter. There may be evidence of the copyist's natural handwriting style in the simulation and other differences from the genuine writing, especially where an attempt is made to write fluently.

Tracing is the second major method of simulating handwriting, with the appropriate letters and words drawn from a pool of writing available to the forger. This method, although it can be effective in suppressing the handwriting style of the copyist, almost inevitably still has the problem of lack of fluency in the writing. Even if at first glance the writing produced appears to be pictorially similar to the genuine writing, it will almost certainly be slowly completed with many stops and pauses in the ink line. Tracing methods can also leave evidence on the questioned document, depending on the method used. For example, traced guide lines in the form of pencil transfers, carbon copy writing and indented impressions can also usually be detected by the techniques available to the forensic document examiner. Should the specimen document from which the simulation has been made become available, this may also bear such evidence in terms of matching of particular letter or word forms, or indentations of the questioned writing.

Signatures contain many fewer letter forms than handwriting but present the same problems to the person attempting the simulation. A further difficulty arises with elaborate signatures as it can be difficult to determine directions and sequence of strokes. As with

handwriting, it is extremely difficult to maintain both fluency and the correct forms of the signature components. Where the copyist attempts a highly fluent simulation, evidence of the simulation process is usually present in the form of substantial departures from normal letter constructions. In the case of many traced or freehand simulated signatures, as only one or two signatures are often used as models, the signatures may be far more similar to each other than would be expected of a group of genuine signatures. This is especially true for tracings.

Factors Influencing Handwriting

Many factors can affect a person's handwriting, and no one person writes so consistently that each letter form is exactly the same. However, the relative method of construction, letter proportions etc. remain consistent within a small range of variation, even if the writing is completed on an uneven surface, at speed or under some other stress. More significant variations in writing style are caused by such factors as age, injuries, illness (mental or physical) with handwriting showing a reduced speed, tremor in the form of erratic impulse movements and there may also be misplaced or poorly joined strokes. Attempted imitation of this type of writing sometimes shows a high frequency tremor which may not be consistent with the tremor of the genuine writer, inconsistent letter constructions or careful retouching of strokes which exceeds the skill of the genuine writer. At times the use of medication can improve a person's handwriting for a limited period of time, and this is considered by the document examiner.

The only scientific method of determining whether a piece of handwriting or a signature has been written by a particular person whose handwriting may have been affected by such factors as described above, is to obtain as much comparable writing as possible, written when the factors applied. This normally means obtaining handwriting or signature specimens written as close as possible to the date of the questioned material.

Determination of the genuineness or otherwise of handwriting and/or signatures of the infirm can be among the most difficult examinations which the forensic document examiner undertakes, especially when adequate specimen signatures or handwriting reflecting the infirmity are lacking. Nevertheless, useful, if not always certain, determinations can be made by the standard methods of careful examination and comparison with the available specimens, looking in particular for similarities or differences in the more subtle features of the handwriting and/or signature.

Examination from Reproduction Documents

Reproductions of handwritten documents, in the form of photocopies, fax copies and computer-imaged reproductions are often submitted to the document examiner either as questioned or specimen writings. Of necessity, examination from a reproduction provides reduced information. For example, microscopic examination is of little value, except in determining the nature of the copy, since the fine details of writings are lost in the resolution provided by the copy. Nevertheless, useful handwriting comparisons can be made from examination of reproduced documents, particularly good clear modern photocopies, although the examination is restricted mainly to the grosser pictorial features of the writing. Suitably qualified conclusions are usually expressed, along with warnings that it cannot necessarily be assumed that a true reproduction of an original document has been examined as photocopy or computer manipulation may sometimes be accomplished without leaving evidence of the manipulation in the resulting reproduction.

Other Examinations

Document examination encompasses much more than comparison of handwriting and signatures. Document examiners also consider the inks and printing seen on the paper, the paper itself, folding, stamps and seals and writing impressions. All of these, together with the handwriting and signatures, go towards proving the provenance of any particular questioned document.

See also: **Document Analysis**: Analytical Methods; Forgery/Counterfeits; Ink Analysis; Document Dating.

Further Reading

Conway JVP (1959) *Evidential Documents*. Springfield, IL: Charles C Thomas.

Ellen D (1997) *The Scientific Examination of Documents*, 2nd edn. London: Taylor & Francis.

Harrison WR (1966) *Suspect Documents*. 2nd impression with supplement. London: Sweet & Maxwell.

Hilton O (1982) *Scientific Examination of Questioned Documents*, revised edn. New York: Elsevier Science.

Haber RA and Headrick AM (1959) *Handwriting Identification: Facts and Fundamentals*. Boca Rota, FL: CRC Press LLC.

Osborn AS (1922) *The Problem of Proof*. Albany, NY: Matthew Bender.

Osborn AS (1929) *Questioned Documents*, 2nd edn. Albany, NY: Boyd Printing.

Osborn AS and Osborn AD (1946) *Questioned Document Problems*, 2nd edn. Albany, NY: Boyd Printing.

Ink Analysis

R L Brunelle, Brunelle Forensic Laboratories, Fredericksburg, VA, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0479

Introduction

Chemical and physical analysis of inks on questioned documents provides valuable information regarding their authenticity. Comparison of these chemical and physical properties of two or more inks can determine: (1) if the inks were made by the same manufacturer; (2) in some cases, whether the inks are products of the same production batch; and (3) the first production date of the specific ink formulation involved. When dating tags are detected, it is possible to determine the actual year or years when the ink was manufactured. Dating tags are unique chemicals that have been added to ball-point inks by some ink companies as a way to determine the year the ink was made.

Relative age comparison tests performed on inks of the same formula and written on the same type of paper with the same storage conditions (performed by measuring changing solubility properties of inks) can estimate how long inks have been written on paper. This is done by: (1) comparing the rates and extents of extraction of questioned and known dated inks in organic solvents by thin-layer chromatography (TLC) densitometry; (2) comparing changes in dye concentrations by TLC and TLC densitometry; and (3) comparing the volatile ink components by gas chromatography–mass spectrometry (GC-MS). In cases where known dated writings are not available for comparison with questioned inks, accelerated aging (heating the ink to induce aging of the ink) can sometimes be used to estimate the age of ink using any or all of the above described techniques. Iron-based inks can be dated by measuring the migration of iron along the fibers of the paper by Scanning auger microscopy.

This article describes state of the art procedures for the chemical and physical comparison, identification and dating of inks on questioned documents.

Composition of Major Types of Writing Inks

Knowledge of the composition of inks is necessary to understand the reasons for the various methods used to analyze inks. Also, knowledge of the first production date for each type of ink or certain ingredients in the inks is useful for dating inks.

Carbon (India) ink

In its simplist form carbon inks consist of amorphous carbon shaped into a solid cake with glue. It is made into a liquid for writing by grinding the cake and suspending the particles in a water–glue medium. A pigmented dye may be used to improve the color. Liquid carbon inks are also commercially available. In the liquid carbon inks shellac and borax are used in place of animal glue and a wetting agent is added to aid in the mixing of the shellac and carbon. Carbon inks are insoluble in water, very stable and are not decomposed by air, light, heat, moisture or microbiological organisms. This class of ink has been available for more than 2000 years.

Fountain pen inks

There are two types of fountain pen inks: (1) iron-gallotannate type and (2) aqueous solutions of synthetic dyes. Modern inks of type (2) contain synthetic blue dyes to provide an immediate blue color to the ink which gradually turns black after oxidation on paper. This explains the origin of the name blue–black fountain pen ink. This class of ink is also very stable. This ink is insoluble in water and cannot be effectively erased by abrasion. The most popular fountain pen ink (developed in the 1950s) consists of an aqueous solution of synthetic dyes. These inks are bright and attractive in color, but they are not nearly as stable as the carbon or blue–black inks. Some of the synthetic dyes used fade and are soluble in water. The most modern inks of this type contain pigmented dyes, such as copper phthalocyanine (introduced in about 1953) which makes these inks much more permanent.

Ballpoint inks

The ballpoint pen was developed in Europe about 1939 and was initially distributed in Argentina about 1943. In 1946, several million Reynolds ballpoint pens reached the market in the United States.

Ballpoint inks consist of synthetic dyes (sometimes carbon or graphite is also added for permanence) in various glycol solvents or benzyl alcohol. The dyes in ballpoint inks can consist of up to 50% of the total formulation. Several other ingredients are usually added to the ink to impart specific characteristics. These ingredients consist of fatty acids, resins, surface active agents, corrosion control ingredients and

viscosity adjustors. The fatty acids (oleic is the most common) act as lubricants to the ball of the pen and they also help the starting characteristics of the ball point.

Ballpoint inks made before about 1950 used oil-based solvents such as mineral oil, linseed oil, recinoleic acid, methyl and ethyl esters of recinoleic acid, glycerin monoricinoleate, coconut fatty acids, sorbital derivatives, and plasticizers such as tricresylphosphate. Modern ballpoint inks (post-1950) are referred to as glycol-based inks, because of the common use of ethylene glycol or glycol derivatives as a solvent for the dyes. Benzyl alcohol is also commonly used as the vehicle (solvent) by some ink manufacturers. Chelated dyes (introduced commercially around 1953) are stable to light. Red, green, yellow and other colored chelated dyes are now used for various colored ballpoint inks.

Pressurized ballpoint inks were developed about 1968. These pens contain a pressurized feed system instead of gravity flow. The physical characteristics of these inks are quite different from the standard glycol based ballpoint inks. The composition is basically the same, but this ink does not become fluid until disturbed by the rotation of the ball point in the socket. Cartridges containing this ink are under the pressure of nitrogen or some other inert gas. The positive pressure on the ink allows the pen to write in all positions and in a vacuum. These pens are used by astronauts during space travel.

Rolling ball marker inks

Rolling ball marker inks were introduced in Japan in about 1968 and shortly thereafter in the United States. These inks are water based and usually contain organic liquids such as glycols and formamide to retard the drying of the ball point. The dyes in these inks are water soluble or acidic dye salts. The light fastness of these dyes range from good for the metalized acid dyes to poor for some of the basic dye salts. Water fastness is usually poor, except that some of these dyes have an affinity for cellulose fibers in paper which produces a degree of water fastness. Water-resistant rolling ball marker inks are also available. These inks are totally insoluble in water and can only be dissolved in strong organic solvents, such as pyridine or dimethylsulfoxide (DMSO).

Fiber or porous tip pen inks

This class of inks was developed in Japan about 1962 and in the United States about 1965. Fiber tip inks are usually water or xylene based and contain dyes and additives similar to those in rolling ball marker inks and fountain pen inks. The water-based inks are obviously water soluble, whereas the xylene-based inks are water resistant and can only be dissolved with strong organic solvents. Formamide or glycol solvents are essential ingredients in fiber tip inks to keep the fiber tip from drying out. Fiber tip inks that contain metalized dyes are light fast.

Gel-pen inks

The most recent development in the writing instrument industry is the introduction of the gel-pen by the Japanese. Four brands of gel-pen inks have been introduced: (1) the Uniball Signo by Mitsubishi; (2) the Zebra J-5; (3) the Pentel Hybrid; and (4) the Sakura Gelly Roll pen. These pens have been marketed by the Japanese since the mid-1980s and a limited supply of the pens was sold in the United States about 1993. Two US manufacturers are now producing these pens.

Gel inks contain completely insoluble colored pigments rather than organic dyes. Writing with this ink is very similar to the appearance of the writing with a ballpoint pen. This ink, which is water based, is a gel and not a liquid. It is insoluble both in water and strong organic solvents. This physical property makes it impossible to analyze (by traditional methods) for the purpose of comparing two or more inks of this type.

Ink Comparisons and Identifications

Inks are usually examined for three reasons:

1. To compare two or more ink entries to determine similarities or differences in inks which can provide information concerning whether entries have been added or altered.
2. To determine if two or more entries were written with the same formula and batch of ink, thus providing a lead as to whether certain entries could have been written with the same pen.
3. To date ink entries to determine whether documents have been back-dated. This section deals with the first two reasons for analyzing inks.

Nondestructive methods of comparison should be carried out first, because chemical analysis causes minor damage to the document by removing ink samples for analysis. Typically, the nondestructive methods include: (1) a visual and microscopic examination of the writing to assess its color and the type of pen used; (2) infrared reflectance and luminescence

examinations to determine whether the inks reflect or absorb infrared light and whether the inks luminesce; and (3) viewing the inks under long- and shortwave ultraviolet light to determine if the inks are fluorescent under these wavelengths of light. Often these techniques are sufficient to determine if two or more inks are different. However, if these techniques fail to detect any differences in the inks, then further chemical analysis is necessary to determine if the inks being compared really have the same formula.

The most widely used technique for comparing and identifying inks is TLC. This technique separates the dyes in the ink and the invisible organic components in the ink. This allows a direct comparison of the composition of inks being examined on the same TLC plate. To determine the relative concentrations of dyes present in the ink, the dyes separated on the TLC plate are scanned in a TLC scanning densitometer. The method is fast, reliable and inexpensive. High performance liquid chromatography (HPLC) has also been used for comparing inks with some success. Gas chromatography-mass spectrometry (GC-MS) is a very useful technique but the equipment is expensive.

Method of chemical analysis

Equipment, materials and solvents

- Merck HPTLC plates (silica gel without fluorescent indicator). The plates should be activated at 100°C for 15 min before use.
- TLC scanning densitometer
- Reagent grade pyridine, ethyl acetate, 1-butanol, ethanol, benzyl alcohol, DMSO, and water
- 1 dram (1.8 g) glass vials with screw caps
- 10 μl and 4 μl disposable micropipettes
- TLC glass developing chamber to accommodate standard 4 in × 8 in (10 cm × 20 cm) TLC plates with cover
- 20 guage syringe needle and plunger (the point of the needle must be filed so that the point is flat)
- 10 μl and 20 μl automatic pipettes
- temperature controlled oven

Procedure

- Using the syringe needle and plunger, punch out about 10 plugs of ink from the written line.
- Place the plugs in the glass vial and add 1–2 drops of the appropriate solvent to the vial to dissolve the ink (usually pyridine for ballpoint ink and ethanol and water (1:1) for nonballpoint inks. Water resistant nonballpoint inks may require using pyridine or DMSO). Allow 15 min for the ink to dissolve.
- Note and record the color of the ink in solution and then spot the ink on to the TLC plate using the 10 μl micropipette. Keep the spots small by spotting intermittently and allowing the spots to dry between each spotting.
- Repeat the above for all ink samples to be compared. Up to about 20 samples can be spotted on the same TLC plate. Be sure to analyze a sample of the paper with no ink as a control.
- Place the TLC plate with the spotted inks in a temperature-controlled oven for approximately 10 min at 80°C. Allow the plate to cool to room temperature then place the plate in the developing chamber using a solvent system of ethyl acetate:ethanol:water (70:35:30 by vol.). The solvent system should be allowed to equilibrate in the developing chamber for at least 15 min.
- Allow the TLC to develop for 15 min, then remove it from the chamber and dry in the oven for approximately 15 min at 80°C.
- View the developed TLC visually and under ultraviolet light to determine which inks match in terms of the dyes and fluorescent components present.
- Scan the plate in the scanning TLC densitometer to measure the relative concentrations of the dyes present in the inks. The dyes are scanned at 585 nm for blue and black inks if a spectrometer type densitometer is used. Video densitometers see all spots in shades of black and therefore no wavelength setting is needed for this instrument. (If the above solvent system did not adequately separate the dyes in the ink for accurate densitometer readings, repeat the tests using 1-butanol:ethanol:water (50:10:15, by vol.).
- Compare the relative concentrations of the dyes present in the various inks. Failure at this point to detect any significant differences among the inks compared justifies a conclusion that all inks are consistent with being of the same formulation. This statement is based on the finding of no significant differences in the nondestructive tests and the chemical analysis. It should be noted that complete identification of an ink is not possible, because not all of the original ingredients in ink are present in ink dried on paper.
- To identify the manufacturer and specific formulation of questioned inks, standard inks of known manufacture and formulation must be analyzed simultaneous with the questioned inks using the same procedures described above. To do this, however, requires access to a complete and comprehensive collection of standard inks and an analytical method that distinguishes each standard. The strength of any identification is only as strong as

the completeness of the standard ink reference collection and the ability to identify its inks.
- If the ink is properly identified, it is possible to determine from the manufacturer when that specific formulation of ink was first made. This may determine if a document was backdated.

Although the above procedures are the most commonly used and have withstood the test of the courts for the comparison and identification of inks, other methods are available. For example, gas chromatography (GC) and GC-MS can be used to detect any volatile organic ingredients that might be present in the inks. HPLC can be used to detect volatile and nonvolatile components. Electron microscopy can be used to distinguish carbon from graphite, when these are present in inks. Time and the amount of ink sample available for analysis usually make the use of these techniques impractical.

Dating of Inks

As mentioned earlier in this article, there is a huge demand for the dating of inks on questioned documents. Any time during an investigation when there is some question about the date of preparation of a document, an ink dating chemist is needed. Over the past 30 years, the ability to perform these examinations has become widely known and recognized among forensic scientists, document examiners and attorneys throughout the world. The ink dating procedures that will be described have passed the Frye and Daubert tests on numerous occasions and are therefore routinely accepted in US courts. Testimony has also been admitted using these techniques in Israel and Australia.

First date of production method

After the ink is uniquely/positively identified, the first date of production of that ink or certain ingredients in the ink is determined from the manufacturer of that specific ink formulation. If the ink was not made until after the date of the document, then it can be concluded that the document was backdated. If the ink was available on the date of the document, then the document could have been written on that date.

Ink tag method

If an ink tag is identified in an ink, it is possible to determine the actual year or years when an ink was made. Tags have been added to some ballpoint inks by the Formulab Company since before 1970; however, the use of tags in their inks was discontinued in June 1994. Since the tags are considered proprietary information by Formulab, no further information about the tags can be reported here. Formulab should be contacted directly, if this information is needed.

Ink dating tags are detected and identified by TLC using a solvent system of chlorobenzene and ethyl acetate (5:1, v/v). Standard samples of the tags should be run simultaneously on the same TLC plate as the questioned inks. The tags, if present, are viewed under longwave ultraviolet light and the R_F values of the tags present in questioned inks are compared with the R_F values of the standard tags. The dates the various tags were used must be obtained from Formulab.

Relative age comparison methods

Dating inks by this procedure is based on the scientifically proven premise that as ink ages on paper, there are corresponding changes in the solubility properties of the inks. Therefore, by comparing the solubility or extraction properties of questioned inks with known dated inks of the same formula on the same type of paper and stored under the same conditions, it becomes possible to estimate how long the ink has been written on the document. Two or more inks of the same formulation can be compared without known dated writings to determine whether the writings were made at the same or different times. This is only true if the inks being compared are still aging (drying), because after the ink has aged out (completely dry), no differences in solubility properties are expected, even if the inks were written at different times. Typically inks will become totally dry (as measured by these procedures) within 6 years; some inks become dry in less than 6 years.

When two or more matching inks are compared without known dated writings, it is still possible to determine the sequence in which the inks were written. This again requires knowing that the inks are still aging and also knowing how the inks age. For example, some inks extract faster and more completely in organic solvents as the ink ages; whereas, others extract more slowly and less completely as they age. To determine which way the ink ages, a sample of the ink is heated at 100°C for 30 min. The rate and extent of extraction of this heated sample into an organic solvent is compared with an unheated sample of the same ink to determine if the heated (totally aged) sample extracted faster and more completely than the unheated sample, or vice versa.

R-Ratio (rate of extraction) method and percent (extent) extraction method

- Using the syringe and plunger, remove 10–15 plugs of ink and paper and place them into 1 dram glass vials. Cap and label the vial with the sample number. Repeat for each sample to be analyzed.
- Set the timer to 10 min.
- Using the automatic 20 µl pipette, add 20 µl of a weak solvent to the vial containing the ink sample and start the timer immediately. (For almost all ballpoint inks, 1-butanol is a good weak solvent.)
- Stir by rotating the vial containing the ink and weak solvent immediately after adding the weak solvent and just before each aliquot is removed for spotting.
- Spot 4 µl aliquots of dissolved ink in one continuous application on a TLC plate at 0.5, 1.5, 3 and 10 min intervals. Place these spots side by side at one end of the plate approximately 1 cm apart. (It may be necessary to use tweezers to remove the pipette from the vial.) Note: If a nonballpoint ink is being analyzed, it may be necessary to spot the 4 µl aliquots intermittently to prevent the spot from getting too large. The spot should be no larger than 0.3 cm in diameter
- Repeat the above procedures for each sample to be analyzed.
- Evaporate the solvent remaining in the vials in an oven at 80°C (about 15 min).
- Remove the vials from the oven and allow them to cool to room temperature.
- Using the automatic pipette, add 10 µl of a strong solvent to each vial and allow to extract for 15 min. (Benzyl alcohol is the solvent of choice for ballpoint inks and some nonballpoint inks. Some nonballpoint inks may require using ethanol:water (1:1) or DMSO for water resistant nonballpoint inks.)
- Spot 4 µl of the ink extracted with the strong solvent adjacent to the weak solvent spots. (If benzyl alcohol is used for the strong solvent, spot in one continuous application of the pipette to the plate. If pyridine is used, spot intermittently to keep the spot from getting too large.)
- Repeat the above steps for each sample.
- Dry the spots on the TLC plate at 80°C for about 15 min.
- Remove the plate from the oven and allow to cool to room temperature.
- Scan the plate in the scanning TLC densitometer along the path of the four weak solvent spots and the one strong solvent spot and read the relative concentrations of the five spots.
- Repeat the scan described above for each sample.
- Calculate the various R-ratios for each sample by letting the percent of ink extracted in the weak solvent at 10 min equal one. Then calculate the R-ratios for each time interval of 0.5, 1.5, 3 and 10 min. This gives a normalized curve.
- To obtain R-ratio curves, plot R-ratios vs. time of extraction (**Fig. 1**). Since all samples are being compared in the same manner, it is not necessary to correct for volume changes caused by successive aliquots removed from the vials.
- Compare the R-ratio curves of all inks tested of the same formulation. To estimate the age of the questioned inks, compare the R-ratio curves of the questioned inks with known dated inks.
- Calculate the percentage or extent of ink extracted in the weak solvent at the various time intervals, by dividing the reading for each weak solvent spot by the total amount of ink extracted in the weak and strong solvent, then multiply by 100. **Figure 2** shows the amount of ink extracted in 10 min.

A simplified percent extraction procedure can be performed by extracting each sample for just 1 min in the weak solvent, then after spotting this 1 min extract, the strong solvent is added directly to the weak solvent remaining in the vial. Then after allowing the strong solvent to extract for 15 min a second aliquot is spotted. This procedure produces just two spots to measure in the densitometer. Although R-ratios cannot be determined by this procedure, accuracy and reproducibility of the percent extraction

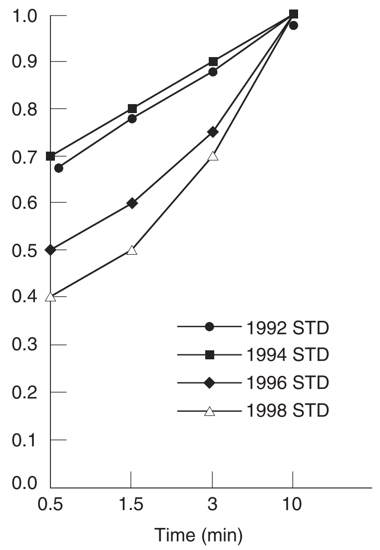


Figure 1 R-Ratio curves (rates of extraction) for Bic black ballpoint inks written in 1992, 1994, 1996 and 1998. The matching curves for the 1992 and 1994 inks means that the ink became totally dry after 4 years, because there was no change in rate of extraction after this time.

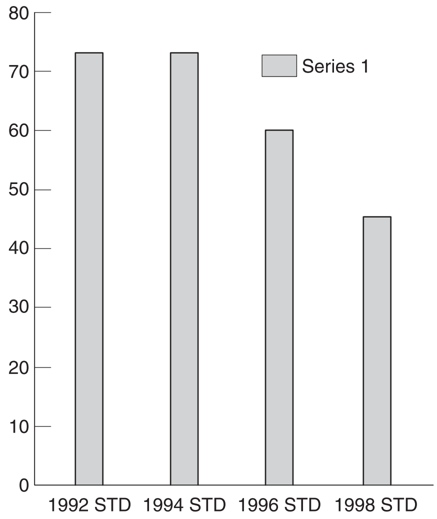


Figure 2 Percent extractions for Bic black ballpoint inks written in 1992, 1994, 1996 and 1998. As with the R-ratios in Fig.1, the age of this ink could only be distinguished between 1994 and 1998. Notice more ink was extracted from the older inks than the newer inks. This is an example of a reverse extraction rate and extent, because most inks extract more from newer inks.

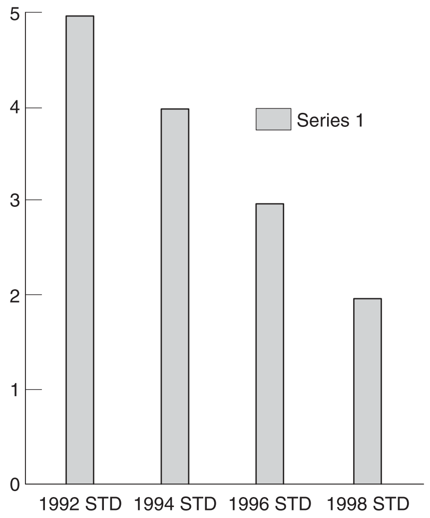


Figure 3 Dye ratios for Bic black ballpoint inks written in 1992, 1994, 1996 and 1998. This method allowed the age of all inks to be distinguished between 1992 and 1998. This is because the relative concentrations of the dyes in this ink changed with age up to 6 years. For this ink the solubility properties as measured by the R-ratio and percent extraction methods did not change after 4 years.

measurements are improved by reducing the number of steps in the procedure and increasing sensitivity. If this procedure is followed, only 10 μl of weak solvent is needed and sample size can be reduced to less than 10 plugs of ink.

Dye ratio method The same plate used for the R-ratio and percent extraction measurements can be used to calculate the various dye ratios for each ink being compared.

- Develop the TLC plate containing all the spots from the R-ratio and percent extraction measurements in a solvent system of ethyl acetate:ethanol: water (70:35:30, by vol.) for 15 min.
- Dry the plate in an oven set at 80°C for about 10 min, then allow the plate to cool to room temperature.
- Scan each sample in the densitometer along the direction of the dyes separated in each sample and from the densitometer readings calculate all possible dye ratios for each sample. For example, divide dye 3 by dye 1, divide dye 3 by dye 2, and divide dye 2 by dye 1. Compare the dye ratios of corresponding pairs of dyes obtained for questioned and known dated inks to estimate the age of the questioned inks (**Fig. 3**).

Since it has been established that these dye ratios change as ink ages, inks with matching dye ratios are consistent with the inks being written at the same time. Inks with dye ratios that do not match generally means that the inks were written at different times, unless one ink had an unusually large batch variation.

It is important to know that depending on the ink formulation involved and the paper it is on, each of the methods described above may not all have equal ability to discriminate the age of the ink being analyzed. For example, it is not uncommon for one method to detect differences in age, when one or both of the other procedures fail to detect this difference. This fact does not negate the positive results of the one method. Only if the results of one method conflict with the results of another method are the overall results negated.

Accelerated aging In situations where known dated inks are not available for comparison with questioned inks, accelerated aging of a questioned ink can be performed to estimate its age (**Fig. 4**). The measurement procedures are identical to those described for R-ratios, percent extraction and dye ratios. This test involves just one additional step which is to heat a sample of the questioned ink for 30 min at 100°C, allow it to cool and equilibrate with the temperaure and humidity in the room for 1 h and then compare the results of the various measurements (using any or all of the R-ratio, percent extraction and dye ratio methods) with the results obtained from an unheated sample of the same ink.

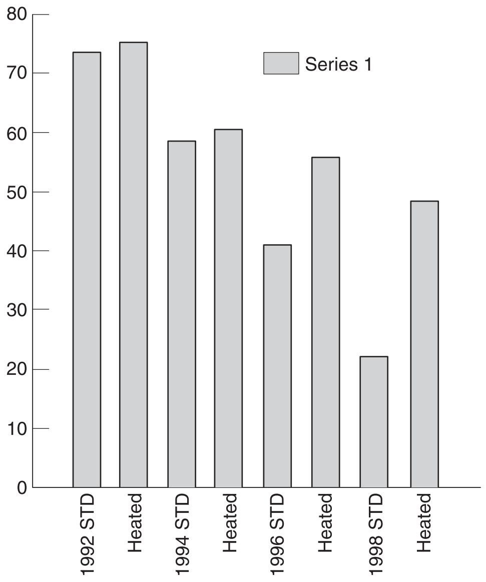


Figure 4 Artificial aging of Bic black ballpoint inks written in 1992, 1994, 1996 and 1998. Notice that the newer inks changed more with heat than the older inks. Also, notice that after 4 years, this ink did not change with heat as measured by the percent extraction method. With knowledge of how long it took for this ink to become completely dry, it is possible to estimate the age of questioned Bic black ballpoint inks by determining how much the extraction properties change with heat.

Significant differences obtained by any one of the methods indicates that the ink is still drying and is therefore less than 6 years old, since no inks have been found to take longer than 6 years to become completely dry using these methods. If it is known that the specific ink in question takes only 3 years to dry, then it can be concluded that the questioned ink is less than 3 years old. This method can also be used to determine which of two or more inks is newer than the other. This is done by observing which ink changes more with heat; the larger the change caused by heat, the newer the ink. This can only be done when all inks compared consist of the same ink formulation on the same type of paper and stored under the same conditions. This statement applies to all of the relative age comparison techniques described here.

See also: **Document Analysis**: Handwriting; Analytical Methods; Forgery/Counterfeits; Document Dating.

Further Reading

Aginsky VN (1996) Dating and characterizing writing, stamp pad and jet printer inks by gas chromatography/mass spectrometry. *International Journal of Forensic Document Examiners* **2**:103–115.

Brunelle RL (1992) Ink dating – the state of the art. *Journal of Forensic Sciences* **37**: 113–124.

Brunelle RL (1995) A sequential multiple approach to determining the relative age of writing inks. *International Journal of Forensic Document Examiners* **1**:94–98.

Brunelle RL and Cantu AA (1987) A critical evaluation of current ink dating techniques. *Journal of Forensic Sciences* **32**: 1511–1521.

Brunelle RL and Lee H (1989) Determining the relative age of ball-point ink using a single solvent extraction mass-independent approach. *Journal of Forensic Sciences* **34**:1166–1182.

Brunelle RL and Reed RW (1984) *Forensic Examination of Ink and Paper*. Springfield, IL: Charles C Thomas.

Brunell RL and Speckin EJ (1988) Technical report with case studies on the accelerated aging of ball-point inks. *International Journal of Forensic Document Examiners* **4**:240–254.

Cantu AA (1988) Comments on the accelerated aging of inks. *Journal of Forensic Sciences* **33**:744–750.

Cantu AA (1996) A sketch of analytical methods for document dating, Part II. The dynamic approach: determining age dependent analytical profiles. *International Journal of Forensic Document Examiners* **2**:192–208.

Cantu AA and Prough RS (1987) On the relative aging of ink – the solvent extraction technique. *Journal of Forensic Sciences* **32**:1151–1174.

McNeil RJ (1984) Scanning auger microscopy for dating of manuscript inks. In: Lamber JB (ed.) *Archeology Chemistry – III*, pp. 255–269. Advances in Chemistry Series No. 205 Washington, DC: American Chemical Society.

Domestic Violence *see* **Clinical Forensic Medicine**: Recognition of Pattern Injuries in Domestic Violence Victims.

DRUGS OF ABUSE

Contents

Analysis
Antemortem Blood
Body Fluids
Classification, including Commercial Drugs
Designer Drugs
Drugs and Driving
Hair
Postmortem Blood
Urine

Analysis

R Kronstrand and **A W Jones**, Department of Forensic Toxicology, University Hospital, Linköping, Sweden

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0494

Introduction

The analysis of drugs of abuse in biological specimens belongs to a branch of the forensic sciences known as analytical toxicology.

Since the beginning of the twentieth century, several epidemics of drug abuse have occurred, in both eastern and western countries, but the use and abuse of illicit drugs started to become a serious problem in society around the mid-1960s, coinciding with the 'hippy' movement, the Vietnam War and the student revolts taking place in many countries. Drug abuse has now infiltrated other segments of the population, especially the inner cities marked by unemployment, poverty, crime and other kinds of deviant behavior. More recently, the abuse of synthetic or 'designer' drugs has escalated among teenagers and adolescents belonging to the rave culture. These substances include ecstasy (methylenedioxymethamphetamine, MDMA) and its analogues, which are produced at clandestine laboratories and marketed at special dance clubs and on the World Wide Web. Some of the chemical structures of these so called 'smart drugs' have not yet been classified as narcotics. In some countries, a fairly liberal attitude towards the sale and consumption of 'soft drugs', mainly tetrahydrocannabinol (THC) derivatives (cannabis, marijuana, Indian hemp) has become evident.

Legislation prohibiting the manufacture, production, handling, sale and consumption of narcotic drugs for recreational purposes exists in most countries. Indeed, use of drugs in the workplace is now regulated by statute in the USA, which has created the need for large-scale screening of urine samples to identify the presence of psychoactive substances. Much effort has been directed towards monitoring drug abuse within the criminal justice system, that is, among prison inmates and others held in detention for various crimes. Rehabilitation of drug addicts requires regular monitoring of body fluids for illicit drugs as a means to control for relapse. Driving under the influence of drugs is another area where increasing attention is being given by the news media and government authorities charged with improving traffic safety and reducing drug-related deaths on the highway. The qualitative and quantitative identification of an abused substance in a blood sample from the person suspected of drug-impaired driving provides important prosecution evidence. Analyzing drugs and poisons in body fluids is also a fundamental element of postmortem toxicology when the cause of death has to be established, and also in emergency medicine when overdosing with drugs is suspected.

Classification of Abused Drugs

Drugs of abuse are often classified according to their chemical structure and the pharmacological effects they produce on the body, e.g depressants, stimulants, hallucinogens, and whether or not well-defined medicinal uses exist, e.g. as anxiolytics, sedatives, hypnotics or analgesics, or whether the drugs are simply taken for pleasure and to elicit feelings of euphoria. **Table 1** presents examples of both licit (prescription)

Table 1 Broad classification of drugs of abuse according to their main pharmacological effect

Pharmacological effect	*Drug class*	*Examples*
Central nervous system (CNS) depressants	Barbiturates	Butalbital, pentobarbital, phenobarbital
	Benzodiazepines	Diazepam, oxazepam, flunitrazepam, nitrazepam, temazepam
	Opiates	Heroin, morphine, codeine
	Others	Ethanol, fentanyl, dextropropoxyphene
CNS stimulants	Cocaine	Cocaine, 'crack'
	Amphetamines	Methamphetamine, amphetamine
	Designer drugs	DOM, DOB, MBDB, MDA, MDE, MDMA
	Others	Phenmetrazine, cathinone, methcathinone, khat
Hallucinogens	Phenylethylamines	Mescaline
	Tryptamines	Dimethyltryptamine (DMT), psilocybin, psilocin
	Ergot alkaloids	Lysergic acid diethylamide (LSD)
	Cannabinoids	Hashish, marijuana
	Others	Phencyclidine (PCP)

DOB, 2, 5-dimethoxy-4-bromoamphetamine; DOM, 2, 5-dimethoxy-4-methylamphetamine; MBDB, *N*-methyl-1-(3,4-methylenedioxyphenyl)-2-butanamine; MDA, methylenedioxyamphetamine; MDE, methylenedioxyethylamphetamine; MDMA, methylenedioxymethamphetamine.

and illicit drugs classified according to their pharmacological effects.

Biological Specimens

Forensic and analytical laboratories specializing in drugs of abuse testing might receive specimens of whole blood, plasma, urine, saliva, sweat or hair as materials for analysis. In postmortem toxicology, tissue samples, stomach contents, vitreous humor and cerebrospinal fluid are also among the materials submitted for analysis. The analytical methods used are little different if the specimens are obtained from the living or from the dead, but in autopsy work the interpretation of the results requires some special considerations. Translating the concentration of a drug or narcotic measured in a blood sample obtained at autopsy into the concentration present at the time of death or the dose of drug administered is fraught with difficulties. Most forensic toxicologists consider such estimates unreliable and only general guidelines can be given. The special considerations for interpreting concentrations of drugs of abuse in postmortem specimens will not be dealt with in this section of the encyclopedia.

Drugs of abuse can be determined in either blood or urine and the choice between the two specimens will depend on the purpose of the testing. When impairment of performance and behavior resulting from taking drugs is important, e.g. in connection with drug-impaired driving, the analysis of blood or plasma specimens becomes essential. Being under the influence of drugs requires some knowledge about exposure of the brain to the psychoactive substance, and blood or plasma concentrations are therefore the closest approximation. For monitoring and control of abstinence from use of drugs, urine is the preferred fluid because the detection window is appreciably longer compared with that for analysis of blood samples. **Table 2** gives the approximate detection times for

Table 2 Approximate detection times of drugs of abuse and their main metabolites in blood and urine

		Detection time (h)	
Drug	*Metabolite*	*Blood*	*Urine*
Amphetamine	—	72	up to 168[a]
THC	—	<12	—
	Carboxy-THC	48–168	168—to several weeks[b]
Cocaine	—	<12	24
	Bensoylecgonine	24	72
Heroin	—	<0.5	—
	6-acetylmorphine	3	24
	Morphine	24	72–96

[a] Depends on urinary pH.
[b] Depends on intensity of use.

being able to identify various drugs of abuse and their metabolites in blood and urine specimens.

General Procedures

Fig. 1 shows a flow diagram illustrating the sampling and processing of biological specimens for analysis of drugs of abuse, e.g. in workplace drug testing programs.

Preanalytical

Obtaining the appropriate biological specimen is not a trivial matter; when urine is collected, the sampling should be observed because drug addicts might attempt to adulterate the specimen in various ways. An appropriate volume of specimen (>30 ml) is usually collected in a clean dry container without preservatives (e.g sodium fluoride), taking care to label the sample with the donor's name, and the date and time of sampling. Special tamper-proof security tapes are available and should be used to seal the sample container prior to transport and storage in the laboratory. All these safeguards help to ensure the validity of the final analytical result. The chain of custody of biological specimens from the moment of sampling, during transportation to the laboratory and during the analytical procedure must be documented until the specimen is finally stored in a refrigerator, deep-frozen or destroyed. Chain-of-custody aspects are often much discussed and debated in connection with drugs of abuse testing, especially if the results are challenged, which often happens weeks or months after making the initial analysis.

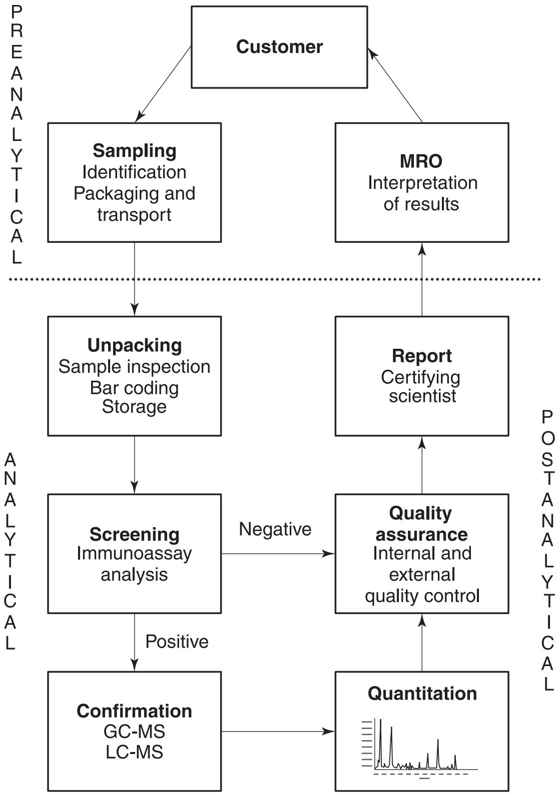


Figure 1 Scheme for obtaining and processing biological samples for analysis of drugs of abuse. MRO, Medical review officer.

In connection with urine drug testing, attempts are sometimes made to manipulate or adulterate the sample, e.g. by dilution with water or other liquids, so that the concentration of the abused drug decreases below the threshold for positive test results. Urine samples might be tampered with *in vitro* after voiding, by adding various chemicals or household products (soap, detergent, lime juice) to make the assay more difficult or impossible to perform, or *in vivo*, by drinking water or alcohol to cause a water-induced diuresis. Determination of the urine creatinine content is one method used to monitor for highly dilute specimens: cutoff concentrations of 0.2–0.5 mg ml^{-1} (1.6–4.0 mmol l^{-1}) are generally used. The temperature of the urine immediately after voiding (32–38°C or 90–100°F), its pH (pH 5–8) and specific gravity (1.003–1.030) are other ways to control for adulterated or otherwise suspicious samples.

Analytical

The most widely used analytical strategy involves making an initial screening analysis to exclude those specimens that do not contain any drugs of abuse above a pre-established threshold concentration. The preliminary screening tests are generally performed on a specimen of urine, whenever available; otherwise a blood sample after protein precipitation can be used for analysis. The drug screening analysis is performed by means of immunoassay techniques. Several classes of drug are usually screened for simultaneously (opiates, cocaine metabolites, amphetamines, cannabinoids as well as lysergic acid diethylamide (LSD) and phencyclidine). Those specimens identified as positive during this screening test are analyzed again using a confirmatory and more specific method of analysis, such as gas chromatography–mass spectrometry (GC-MS).

Postanalytical

Knowledge about the stability of drugs in body fluids during storage becomes an important issue whenever the analytical result has to be verified at some later date. A decrease in concentration of drugs of abuse generally occurs during storage, and data for barbiturates, benzodiazepines, cannabinoids, cocaine, LSD

and phencyclidine have been published. The rate of loss of analyte seems to depend on the physicochemical properties of the drug, especially factors that influence binding to the container, the kind of body fluid analyzed (blood, urine or saliva), the pH of the medium and the storage temperature. Specimens stored deep-frozen will retain the initial drug concentration for longer.

Immunoassay for Drugs of Abuse Testing

Immunoassay technologies were designed to provide a cost-effective way of identifying presumptive positive specimens before embarking on a quantitative confirmatory analysis by a more selective technique. The immunoassay methods for analyzing drugs of abuse were developed primarily in connection with workplace drug testing to meet statutory guidelines. **Table 3** gives the threshold concentrations (cutoff levels) for drugs of abuse in urine as recommended by the Substance Abuse and Mental Health Services Administration (SAMHSA) and the European Union (EU) for drug testing. The SAMHSA guidelines for analysis of opiates were changed during 1998 regarding the cutoff levels for positive results and also the need to identify 6-acetylmorphine as proof of heroin intake.

The basic principle of an immunoassay (IA) involves the reaction between an antibody and an antigen (the drug or its metabolites). The drug molecules in the biological sample compete for binding sites on the antibody with an antigen labeled or marked with either an enzyme (EIA), a fluorescent group (FPIA) or a radioisotope (RIA).

The enzyme multiplied immunoassay technique (EMIT) is by far the most widely used EIA technology for drugs of abuse screening analysis. The EMIT method monitors the change in absorbency at 340 nm when the coenzyme nicotinamide adenine dinucleotide (NAD^+) is converted to its reduced form NADH. The EMIT technology is ideally suited for use with large-scale, fully automated analytical systems that are capable of making thousands of determinations each day (e.g. Hitachi 911). The development of methods for ‘on-site testing’ of urine for drugs of abuse is currently in vogue. Small handheld devices containing the antibody and labeled antigen are available, to which the test material (urine or saliva) is added. A positive or negative test result is obtained within a few minutes of initiating the reaction. To meet the increasing demand for drugs of abuse testing, numerous kits for on-site testing have appeared on the market and these are intended to be used away from the laboratory and without the need for any special training or technical skills on the part of the test operator. Extensive marketing of these on-site drug testing kits has occurred, although their reliability in terms of percent false-positive and false-negative responses differs widely. Some of these drug test kits have been criticized because the performance evaluations were done by trained personnel working in a laboratory environment.

A definite drawback with on-site drugs of abuse testing kits is the need to make a subjective evaluation of the results. Usually this involves identifying and interpreting some kind of color change, which is often not well defined. Recently, however, efforts have been made to develop on-the-spot tests for drugs in saliva. Because the concentrations in saliva seem to correlate well with concentrations in blood, the saliva test might be used at the roadside for screening impaired drivers.

Cross-reactivity of immunoassays

The crossreactivity of immunoassays towards other drugs, metabolites or endogenous substances is an important consideration. The antibodies used in the

Table 3 Guidelines for reporting urinary drug concentrations in connection with workplace drug testing

Screening drug group	*EU ($ng\,ml^{-1}$)*	*SAMHSA ($ng\,ml^{-1}$)*	*Confirmation drug/metabolite*	*EU ($ng\,ml^{-1}$)*	*SAMHSA ($ng\,ml^{-1}$)*
Amphetamines	300	1000	Amphetamine	200	500
			Methamphetamine	200	500[a]
Cannabinoids	50	50	Carboxy-THC	15	15
Cocaine metabolites	300	300	Bensoylecgonine	150	150
Opiates	300	2000[b]	Morphine	200	2000[b]
			Codeine	—	2000[b]
			6-Acetylmorphine	—	10[b]
Phencyclidine	—	25	Phencyclidine	—	25

[a] To be reported as methamphetamine-positive the sample must also contain at least 200 $ng\,ml^{-1}$ amphetamine.
[b] Both screening and confirmation thresholds were increased from 300 to 2000 $ng\,ml^{-1}$ in 1998, with an additional requirement of 6-acetylmorphine determination (at 10 $ng\,ml^{-1}$) in all samples positive for morphine above the new 2000 $ng\,ml^{-1}$ threshold by GC-MS.

reaction exhibit different affinities for different drugs or analogs because the antibodies are always calibrated for a certain drug or metabolite over a specified range of concentrations. Reagents for benzodiazepines might show good sensitivity for nordiazepam and oxazepam, whereas the response to flunitrazepam might be unacceptably low, which can lead to a false-negative result for this sedative drug. The degrees of crossreactivity for each specific assay are usually provided by the manufacturer and these should be carefully scrutinized to understand better the capabilities of the reagent and which drugs or metabolites might interfere with the test.

Methods of Confirmatory Analysis

All specimens identified as positive by an immunoassay screening procedure must be reanalyzed using a more sensitive and selective method. This normally involves use of gas chromatography (GC) or high-performance liquid chromatography (HPLC) in combination with mass spectrometry (MS). Indeed, mass spectrometric methods, preferably with deuterium-labeled internal standards, are considered mandatory in many drugs of abuse testing laboratories. Accordingly, MS is a prerequisite for definitive testing to confirm the positive immunoassay results so that the concentration of abused drug or drugs can be reported with certainty. A summary of analytical techniques used for quantitation is given in **Table 4**.

Confirmatory analysis necessitates taking a fresh aliquot of the blood or urine and extracting the drug from the biological matrix by means of various clean-up procedures. Because many drugs and metabolites excreted in urine are conjugated with glucuronic acid or sulfate, the drug conjugates are often hydrolyzed to liberate the drug and thereby increase its concentration before proceeding with the quantitative analysis. The hydrolysis of glucuronide conjugates is either done by carefully heating the urine with hydrochloric acid or by the use of various enzymes (β-glucuronidase) at a controlled pH. **Table 5** lists various drugs of abuse together with the main metabolites commonly encountered in urine samples sent for analysis to forensic toxicology laboratories.

Clean-up procedures

The methods used to extract drugs from a biological matrix are similar for both blood and urine specimens, although with blood samples one is usually looking for the parent drug as well as metabolites, whereas with urine only the metabolites might be present. No universal procedure exists to extract all drugs from the biological matrix because of the vast differences in physicochemical properties. Owing to the low concentration of an active drug or its metabolites, together with many possible interfering substances, there is a need to choose a particular clean-up procedure that will concentrate the analytes.

The traditional method of extracting a drug and/or its metabolites from a biological matrix (blood, plasma or urine) involved the use of liquid–liquid extraction. This required adjusting the pH of the

Table 4 Analytical techniques used for quantitative analysis of drugs of abuse

Method	*Sensitivity*	*Specificity*	*Expertise/cost*	*Applications for drugs of abuse testing in forensic toxicology*
HPLC-UV	*	*	**	Polar drugs/metabolites, such as benzodiazepines
HPLC-fluorescence	***	**	***	Polar drugs/metabolites at low concentrations, such as LSD
HPLC-EC	***	**	***	Polar drugs/metabolites at low concentrations, such as cannabinoids
HPLC-MS	***	****	*****	Polar drugs that are present at very low concentrations, such as LSD, psilocin
GC-FID	*	*	**	Drugs with oxidizable carbons, present in $\mu g\,ml^{-1}$ concentrations
GC-EC	***	**	***	Halogenated and nitroaromatic drugs, such as benzodiazepines and other drugs that can be derivatized with halogenated reagents
GC-NPD	***	**	***	Drugs at low concentrations containing a nitrogen or phosphorus atom.
GC-MS	***	****	****	Drugs present at low concentrations
GC-MS-MS	****	*****	*****	Drugs at very low concentrations, structural identification of unknowns

* Low; ***** very high.

Table 5 Principal drugs of abuse encountered in forensic toxicology and their main urinary metabolites

Parent drug	*Main metabolites*
Heroin	6-Acetylmorphine, morphine[a]
Morphine	Morphine[a], normorphine
Codeine	Codeine[a], morphine[a], norcodeine
Fentanyl	Norfentanyl
THC[b]	Carboxy-THC[a], 11-hydroxy-THC
Amphetamine	Phenylacetone, 4-hydroxyamphetamine, norephedrine
Methamphetamine	Amphetamine
MDMA (ecstasy)	4-Hydroxy-3-methoxy methamphetamine[a] (HMMA), MDA
MDE	4-Hydroxy-3-methoxy ethylamphetamine[a] (HME), MDA,
MDA	4-Hydroxy-3-methoxy amphetamine[a] (HMA)
Phenmetrazine	4-Hydroxy phenmetrazine[a]
Cocaine	Benzoylecgonine, ecogonine methylester, norcocaine, cocaethylene[c]
LSD	Nor-LSD
Diazepam	Nordiazepam[a], oxazepam[a], temazepam[a]
Oxazepam	Oxazepam[a]
Flunitrazepam	7-Aminoflunitrazepam, 7-acetamidoflunitrazepam
Phenobarbital	Hydroxylated metabolites
Phencyclidine	No major urinary metabolites

[a] Excreted mainly conjugated with glucuronic acid.
[b] THC, Δ^9-tetrahydrocannabinol, the primary psychoactive constituent of cannabis, marijuana, hashish and hemp.
[c] This metabolite, which is pharmacologically active, can be identified if cocaine and ethanol are taken together.

body fluid with a suitable buffer solution so that the substance of interest remained predominantly in unionized (lipid-soluble) form. An organic solvent such as chloroform was then added to the buffered body fluid or tissue homogenate and mixed vigorously, often by mechanical shaking for 30 min or more in an attempt to bring the drug from the aqueous phase into the organic phase. Evaporation of the organic solvent to dryness concentrated the drug prior to making a derivative suitable for chromatographic analysis. Cleaner extracts could be obtained by making a back-extraction from solvent to aqueous phase and again over into the solvent, but this extra step usually gave a lower percentage recovery of the drug.

In recent years, because of time constraints and possible health hazards associated with handling large volumes of organic solvent (e.g. chloroform or ether), other methods of extracting drugs from biological samples have been developed: in particular, solid-phase extraction (SPE). SPE columns or cartridges are commercially available and have become widely used in laboratories specializing in drugs of abuse testing. These disposable SPE cartridges are easy to use and permit the rapid extraction of acidic, basic or neutral drugs from a biological matrix with good recoveries. The columns or cartridges used for SPE are packed with an inert sorbent material such as silica or polymer resins, with attached functional groups, that preferentially bind to the drugs of interest, allowing the impurities to pass through the column. The drug is finally eluted from the column after washing with suitable solvents or buffers. The choice of solvents and the kind of SPE columns depend on the target drug to be analyzed in the biological specimen. Examples of extraction and detection protocols for amphetamine, carboxy-THC, unconjugated opiates and cocaine metabolites are shown in **Table 6**.

Choice of an internal standard

The quantitative analysis of drugs of abuse in body fluids almost always involves the use of an internal standard. Accordingly, the first step in an assay requires adding to the biological specimen a known amount of an analog of the drug of interest before starting the clean-up and quantitative analysis. When GC-MS or HPLC-MS are the methods used, the internal standard of choice is generally a deuterium-labeled analog with a molecular weight usually 3–11 mass units higher than the illicit drug. The chemical properties of the deuterium-labeled analog are practically identical to those of the analyte, and similar mass fragmentation patterns are obtained. The deuterium compounds are slightly more lipophilic than the nonlabeled substances and this might give slightly shorter chromatographic retention times. Adding an internal standard also helps to compensate for incomplete extraction recovery during the clean-up procedures, and at the same time minimizes losses of analyte that occur during transport through the GC column and in the injector port.

Table 6 Confirmatory (verification) methods suitable for analysis of principal drugs of abuse in urine

Assay conditions	*Amphetamines*	*Carboxy-THC*	*Unconjugated opiates*	*Cocaine and metabolites*
Sample volume	0.2 ml urine	1.0 ml urine	1.0 ml urine	1.0 ml urine
Internal standard	Pentadeuterated analogs	Trideuterated carboxy-THC	Trideuterated analogs	Trideuterated analogs
Pretreatment	None	Hydrolysis KOH 120C° for 10 min	None	None
pH setting	KOH pH 13	Acetic acid pH 4	Phosphate buffer pH 6.1	Phosphate buffer pH 6.1
Extraction	Isooctane	Bond Elute THC SPE	Bond Elute Certify SPE	Bond Elute Certify SPE
Derivatization	TFAA	PFPA/PFPOH (4:1)	Acetonitrile: BSTFA 1% TMCS (3:1)	PFPA/PFPOH (4:1)
GC column	30m * 0.25 mm HP 5MS	30m * 0.25 mm HP 5MS	30m * 0.25 mm HP 5MS	30m * 0.25 mm HP 5MS
SIM ions	Amphetamine 140, 118, 91 Methamphetamine 154, 118, 91 Phenmetrazine 273, 167 MDA 275, 140, 162 MDMA 289, 154, 162 MDE 303, 168, 162	Carboxy-THC 622, 607, 459, 445	6-acetylmorphine 399, 340, 287 Morphine 429, 414, 401 Ethylmorphine 385, 357, 234 Codeine 371, 343, 234	Cocaine 303, 272, 182 Bensoylecgonine 361, 346, 240
LOQ	200 ng ml^{-1}	5 ng ml^{-1}	20 ng ml^{-1}	50 ng ml^{-1}

KOH, potassium hydroxide; SPE, solid-phase extraction; TFAA, trifluoracetic acid anhydride; PFPA, pentafluorpropionic acid anhydride; PFPOH, pentafluoropropanol; BSTFA, *N,O*-bis(trimethylsilyl)trifluoroacetamide; TMCS, trimethylchlorosilane; SIM, selected ion monitoring; LOQ, limit of quantitation (10 times the signal-to-noise ratio).

Derivatization

Derivatization of an analyte entails converting the molecule into a structure more suitable for chromatography and/or detection. It also assists in: (1) enhancing the analyte's thermal stability and volatility; (2) improving the analyte's chromatographic characteristics; (3) increasing the sensitivity of the assay; and (4) ensuring optimal mass spectrometric characteristics of the analyte.

Many drugs and their metabolites contain functional groups with active hydrogen atoms, for example amino groups, alcohol groups and carboxylic acid groups. Active hydrogens tend to enhance polarity and facilitate hydrogen bonding, which negatively influences volatility, thermal stability and reactivity. After derivatization, the drug passes more easily through the GC injector without being thermally degraded and without binding to the chromatographic column material.

Although derivatization is an easy task to perform, care is necessary to optimize the conditions. The extract should not be heated to high temperatures for prolonged periods because this can produce interfering substances and lower recovery of the analyte. The obvious way to choose a derivative is to inspect the molecular structure and functional groups present in the drug, and then consider which analytical techniques are suitable. This will often determine the strategy for preparing a suitable derivative. Derivatization requires additional steps in the pretreatment of the sample and therefore should be easy to perform, with the potential for automation.

Table 7 gives examples of the principal derivatizing agents applied to the analysis of some common drugs of abuse. For GC and GC-MS the usual procedures are silylation, alkylation and acylation reactions. Silyl derivatives tend to be easily hydrolyzed and the extracts should therefore be protected from moisture during the reaction. This can be achieved by dissolving the sample in a solvent containing a small amount of the derivatizing agent, or even by injecting the reaction mixture directly into the GC column. Alkylation includes replacement of an active hydrogen with an aromatic or aliphatic group, e.g. a methyl group to produce a volatile derivative with good stability. Acylation involves the conversion of alcohols and amines into esters and amides, respectively. The most common acylating reagents are anhydrides containing halogens, preferably fluorine, such as pentafluoropropionylanhydride (PFPA). Perfluoroacylation improves the sensitivity of GC analysis, especially when an electron capture detector is used for analysis, and at the same time the sensitivity and selectivity of mass spectrometric detection is also improved.

Table 7 Derivatization reagents and applications for confirmatory analysis of drugs of abuse

Functional group	*Drugs of abuse*	*Reagents*	*Derivatization procedure*	*Comment*
Primary amine-NH_2	Amphetamine MDA	TFAA, PFPA, HFBA	Perfluoroacylation	Amines form stable amides with the acetylating reagents and are preferred to silylation
Secondary amine -NH-R	Methamphetamine MDMA, MDE, MBDB	TFAA, PFPA, HFBA	Perfluoroacylation	
Carboxyl -COOH	Benzoylecgonine	PFPA/PFPOH BSTFA, MSTFA	Perfluoroacylation Trimethylsilylation	Carboxylic acids form esters in a two step reaction with PFPA/PFPOH
Carboxyl -COOH and hydroxyl -OH	Carboxy-THC	Methyl-8, TMSH PFAA/PFPOH	Alkylation Perfluoroacylation	Methylation usually produces more stable derivatives and less fragmentation compared to perfluoroacyl derivatives
Hydroxyl -OH	Morphine, codeine	TFAA, PFPA, HFBA, BSTFA, MSTFA	Perfluoroacylation Trimethylsilylation	Acylated hydroxyls have less thermal stability than silylethers and a peakfronting can be seen as a result of decomposition on the column

TFAA, trifluoracetic acid anhydride; PFPA, pentafluorpropionic acid anhydride; PFPOH, pentafluoropropanol; HFBA, heptafluorobutyric acid anhydride; BSTFA, *N,O*-bis(trimethylsilyl)trifluoroacetamide; TMSH, trimethylsulfonium hydroxide; MSTFA, *N*-methyl-*N*-trimethylsilyltrifluoroacetamide; methyl-8, *N,N*-dimethylformamide dimethylacetal.

Gas chromatographic analysis

Gas chromatography is the mainstay of analytical toxicology laboratories because this technique allows separation of the components of a mixture and quantitative analysis in a single step. Separation of the components is based on physical or chemical properties of the drugs, such as their partition coefficients, solubility, polarity and volatility. A carrier gas, which is usually helium or nitrogen, serves to transport the analytes from the injection port through a chromatographic column and into the detector. During this process, the components of a mixture equilibrate between the moving and stationary phases and, depending on the rate of movement through the column, a separation is accomplished. After separation the various components and the amounts present are measured both qualitatively and quantitatively by one of a number of detector systems.

Flame ionization detector (FID) The FID is a universal detector that responds to almost any compound containing carbon and hydrogen atoms. The intensity of response depends on the number of oxidizable carbon atoms in the molecule. After passing through the column mixed with the carrier gas, the analytes are burnt in a hydrogen–air flame. The current produced from the ionized compounds is measured by an electrode situated above the flame. The current is amplified and the response fed to an electronic integrator and recording system. The FID has excellent linearity over several orders of magnitude but the sensitivity is limited to the microgram per milliliter range.

Nitrogen–phosphorus detector (NPD) This detector uses the same principle as the FID but by introducing alkali metal vapors through an electrically heated rubidium bead, in combination with a lower flame temperature, this dramatically increases the response to compounds containing nitrogen and phosphorus atoms. By slight changes in the operating conditions one can achieve phosphorus:carbon response ratios of 50 000:1 or nitrogen:carbon response ratios of 5000:1. The detector is widely used in the toxicological laboratory because many drugs contain nitrogen atoms, but the NPD is also very useful for the analysis of, for example, phosphorylated pesticides.

Electron capture detector (ECD) The ECD contains a radioactive isotope, e.g. ^{63}Ni, that bombards the carrier gas with electrons. Electrons are absorbed by compounds with high electron affinity, such as halogenated compounds or those with nitro- or carbonyl groups. The response is measured as the drop in

current, not as an increase as with the other detectors. Benzodiazepines are readily detected by ECD because they often contain both halogens and nitroaromatic structures. The ECD has been used to analyze many drugs, especially if halogenated derivatives can be prepared.

Mass selective detector (MSD) The MSD is the most universal and still the most selective detector, and mass spectrometric methods are widely used for analysis of drugs of abuse, especially when the structural identity of the compound has prime importance. With the MSD the target compounds are excited by either electron bombardment or by the collision energy from a reagent gas (e.g. ammonia or methane) in an ion source. The excited molecules fragment in ways predicted by their functional groups and chemical structure. This gives a 'fingerprint' of ions that shows the characteristics of the molecule and permits an almost unambiguous identification. With GC-MS methods, both retention time and mass fragments are used to identify the unknown substance by comparison with results from known standards. These standards contain the target drug at appropriate concentrations and are extracted and derivatized in the same way as the authentic sample.

Analysis of Specific Drugs

Amphetamines

Amphetamines are central nervous stimulants with a chemical structure related to biogenic amines derived from phenylisopropylamine. Different side chains are possible and various groups might be joined to the aromatic ring, e.g. ecstasy (MDMA). Within this class of drugs amphetamine, methamphetamine, MDA, MDMA and MDE are widely abused. Amines are readily extracted from aqueous solutions, such as urine, at an alkaline pH. The extraction of amphetamines with organic solvent (e.g. isooctane, ethyl acetate or butyl chloride) followed by derivatization with a perfluoroacyl reagent (e.g. TFAA, PFPA or HFBA) are commonly used (Tables 6 and 7). The derivatives can then be analyzed by GC-MS in full scan or selected ion monitoring (SIM) mode, with at least three prominent fragments being used for identification and quantitative analysis. Derivatization of the analyte is necessary before embarking on GC-MS analysis of amphetamines because of their low molecular weights and the fact that they undergo alpha-cleavage thus producing base peaks with mass-to-charge ratio (m/z) 44, 58 or 72, depending on the particular N-substitution. This low mass range means that background signals and 'noise' are troublesome and numerous endogenous compounds might interfere with the analysis. Increasing the molecular weight by derivatization before analysis yields m/z fragmentation patterns with less background noise, and a more sensitive and specific identification is possible. Many laboratories have developed confirmation methods dedicated to rapid and reliable analysis of amphetamine and methamphetamine because they represent the most commonly abused drugs in many countries. Also the enantiomers of amphetamine and methamphetamine can be separated by GC-MS with the aid of chiral chromatographic columns, or chiral derivatizing reagents. Accordingly it is possible to distinguish the *d*- and *l*-forms of these amines, which might be necessary to rule out legal use of some over-the-counter drugs such as Vick's inhaler which contains the *l*-isomer of methamphetamine.

Fig. 2 shows a chromatogram obtained from the analysis of a control sample containing amphetamine

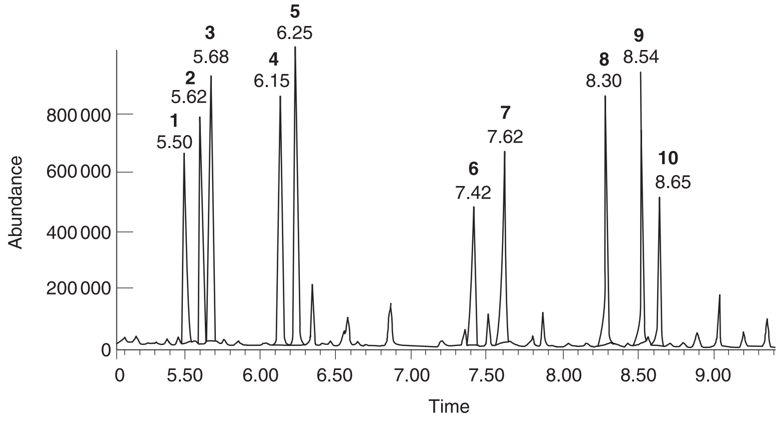


Figure 2 Total ion chromatogram (TIC) from the analysis of a standard sample containing amphetamine and related drugs. 1, Amphetamine; 2, phentermine; 3, norephedrine; 4, ephedrine; 5, methamphetamine; 6, phenmetrazine; 7, methylenedioxyamphetamine; 8, methylenedioxymethamphetamine; 9, methylenedioxyethylamphetamine; 10, methylenedioxyphenyl-2-butanamine.

and related compounds as their perfluoracylated derivatives. The use of this broader analysis is popular in many forensic toxicology laboratories when the preliminary immunoassay screening analysis yields positive results for amphetamines. The confirmation analysis determines exactly which drug or drugs are present that might account for the positive screening test result. Because of the increasing number of phenylethylamines currently being abused, there is a continuous need for development in forensic urine drug testing methods.

Marijuana

The main pharmacologically active substance in preparations from *Cannabis sativa* is tetrahydrocannabinol (THC). There are many metabolites of THC, but most analytical methods have been developed to determine carboxy-THC, the main urinary metabolite. Carboxy-THC is excreted as the glucuronide in urine and, before extraction, a hydrolysis is performed, usually by heating with sodium hydroxide to cleave the glucuronide. Extraction of carboxy-THC can be done either with solvent–solvent extraction at acidic pH or using SPE. Because carboxy-THC is an acid, liquid chromatography became a popular method for separation followed by electrochemical detection for quantitative analysis and HPLC methods gave enough sensitivity to detect drug concentrations in the nanogram per milliliter range. More recent demands for mass spectral identification of drugs of abuse in body fluids made it necessary to change from HPLC to GC-MS for the confirmation of carboxy-THC. The successful GC separation of carboxy-THC requires derivatization of the functional groups and several approaches have been used. Two of the most common derivatization methods are methylation or perfluoroacylation of the alcohol and carboxylic acid groups. Methylation gives rise to few ions with high intensity, whereas perfluoroacylation gives more ions, but these are in the high mass range where the background noise is low, as shown in **Fig. 3**. Thus, both derivatizing procedures result in very sensitive analytical methods for determination of carboxy-THC in body fluids.

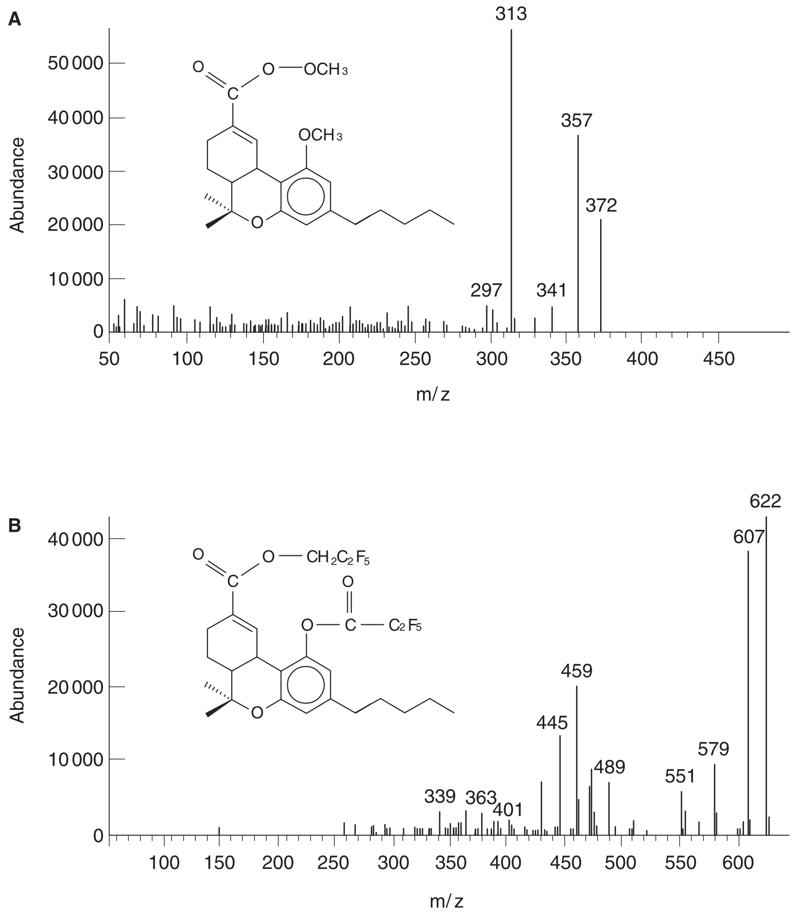


Figure 3 Mass spectrum of **(A)** methylated carboxy-THC using trimethylsulfonium hydroxide (TMSH) as derivatizing agent; and **(B)** perfluoroacylated carboxy-THC using pentafluoropropionic acid anhydride/pentafluoropropanol (PFPA/PFPOH) as derivatizing agents.

Opiates

Analysis of morphine and codeine in urine is readily done with GC-MS methods but special considerations are needed to distinguish which drug has been taken because morphine is a metabolite of codeine (Table 5). Both morphine and codeine are mainly excreted in urine as their glucuronides, and hydrolysis is usually performed before extraction and analysis. Enzymatic (β-glucuronidase) or acid (HCl) hydrolysis can be used to cleave the glucuronide, and recoveries appear to be more effective and reproducible after acid hydrolysis. Because the ratios of total morphine and total codeine are used to distinguish between heroin or codeine intake, the efficiency and reproducibility of the hydrolysis reaction is crucial for correct interpretation.

In addition to evaluating morphine:codeine ratios, one can determine 6-acetylmorphine, which is considered a specific metabolite of heroin, together with the unconjugated morphine and codeine. At a $10\,\mathrm{ng\,ml^{-1}}$ cutoff level for 6-acetylmorphine it may be possible to prove use of heroin if this occurred within about 24 h before sampling the urine.

Despite some difficulties with the chromatography and quantitative analysis, morphine, codeine and 6-acetylmorphine are usually derivatized with trimethylsilylating agents such as BSTFA or MSTFA. Chromatographic separation of morphine, codeine, norcodeine and other opiates, such as hydromorphone and hydrocodone, has proven difficult in routine casework, and the identification by means of qualifier ions is complicated because some ions are the same for different analytes. **Fig. 4** shows a chromatogram and mass spectra of morphine, codeine, hydromorphone and hydrocodone as their TMS derivatives.

Quantitation can be complicated because silicon (^{28}Si) is a so called A+2 element, containing a naturally occurring isotope with a molecular weight 2 mass units higher (^{30}Si). This isotope can interfere in the quantitation if the analyte is present at high concentrations, and if a trideuterated internal standard is used. This problem becomes more acute during the analysis of morphine, which has two TMS molecules attached, and thus two silicon atoms can contribute to the m/z of the internal standard. A better analytical strategy is perfluoroacylation, which gives derivatives without silicon and also produces m/z values in a higher mass range than TMS derivatives. However, the perfluoroacylated derivatives of opiates are thermally unstable and can degrade on the column. Recent developments and lower costs of LC MS instrumentation will prove attractive to forensic toxicologists because the various glucuronides can also be quantitated along with the parent drugs, offering better grounds for judging impairment and evaluating deaths from heroin/opiate overdose.

Cocaine

Cocaine has a short elimination half-life and is rapidly biotransformed after administration. The main metabolites are benzoylecgonine (BE) and ecgonine methyl ester (EME). Other metabolites such as norcocaine and cocaethylene have been analyzed in both blood and urine, but most laboratories use methods that detect cocaine, BE and EME. Cocaine can easily be

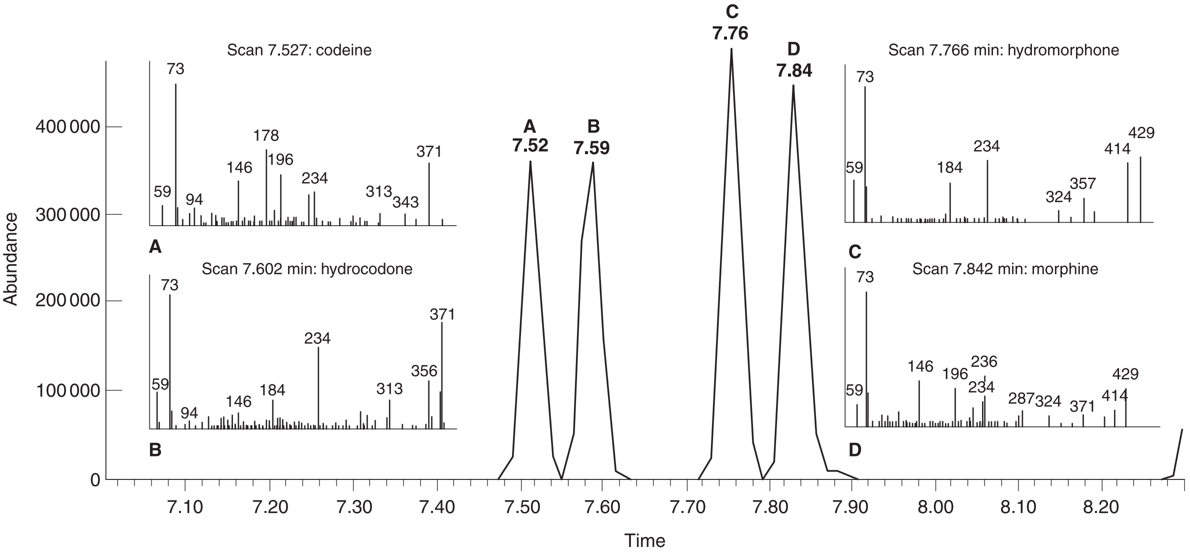


Figure 4 Total ion chromatogram (TIC) and mass spectra from (A) codeine-TMS; (B) hydrocodone-TMS; (C) hydromorphone-2TMS, and (D) morphine-2TMS. Almost baseline separation of the analytes are achieved using a HP-5MS 30 m*0.25 mm inner diameter column. Initial temp 150°C, ramped at $35^\circ\mathrm{C\,min^{-1}}$ to 260°C, then ramped at $5^\circ\mathrm{C\,min^{-1}}$ to 285°C. Incomplete silylation of hydrocodone and hydromorphone as their enols can further complicate the chromatogram.

degraded to BE *in vitro* so precautions are necessary to avoid this happening, particularly when the analytical results are used to estimate the time since administration. BE is a zwitterion, which complicates its extraction from the biological matrix by the possible charge distribution at either the nitrogen atom or the carboxylic acid. However, solid-phase extractions using mixtures of a cation exchanger and a hydrophobic carbon chain have been used with good results. BE contains a free carboxylic acid group which must be derivatized before analysis, whereas cocaine remains unaffected by the derivatization reagent. Both silylating and perfluoroacylating reagents can be used for analysis of cocaine and its metabolites with good results. Moreover, opiates can be extracted under the same conditions as cocaine and BE, which allows identification of heroin and cocaine and their metabolites in the same analytical run.

Hallucinogens

The classical hallucinogens include mescaline, psilocybin and lysergic acid diethylamide (LSD). The two former drugs occur naturally in the peyote cactus and psilocybe mushrooms, respectively, whereas LSD was synthesized by Albert Hoffman in 1943 from ergot alkaloids. The effective doses of LSD and psilocybin are very low and the drugs are extensively metabolized, which creates problems for the forensic toxicologist faced with analysis of these substances in body fluids. Moreover, the drugs are not very stable in body fluids after sampling, which creates difficulties for making reliable quantitative analyses. Psilocybin is rapidly transformed to psilocin, which is excreted in the urine predominantly as a conjugate. Published methods for analysis of psilocybin and psilocin are tedious, involving enzymatic hydrolysis, freeze drying or dialysis. Subnanogram quantitation of LSD in urine was reported by use of HPLC and fluorescent detection after solvent–solvent extraction with diethylether:n-heptane (1:1). Also GC-MS and GC-MS-MS procedures have successfully identified LSD and some of its metabolites at picogram concentrations. The introduction of LC-MS and LC-MS-MS will further increase the potential for qualitative and quantitative analysis of LSD and psilocin in body fluids.

Other abused drugs

A host of other drugs and chemical substances are candidates for abuse and are encountered by forensic toxicologists. In the USA phencyclidine (angel dust) is a leading drug of abuse, although this substance has not been especially prominent in European countries. More recently, γ-hydroxybutyrate (GHB), which is easy to prepare, has received attention as a date-rape drug in some countries. The fast-acting sedative flunitrazepam is also used in date-rape crimes, especially in combination with alcohol. Prescription drugs (sedatives and hypnotics) of the benzodiazepine group are also candidates for abuse but legitimate medical indications make it difficult to legislate against their use. Nevertheless, analytical methods have been published for measuring the concentration of benzodiazepines in body fluids using immunoassay screening techniques and GC-MS with deuterium-labeled internal standards for confirmation.

Concluding Remarks

Developments and innovations in analytical chemistry and instrumentation are helping to keep pace with the escalation in abuse of drugs and other psychoactive substances in modern society. Forensic toxicologists can demonstrate unequivocally that a person has indeed taken a particular drug, provided samples of blood and urine are obtained at appropriate times. Besides a multitude of drug-screening procedures (immunoassays), more sophisticated and time-consuming procedures are necessary for definitive testing. This requires careful extraction of the drug and its metabolites from the biological matrix and performance of a quantitative analysis, along with identification of the chemical structure of the abused substance. In analytical toxicology, mass spectrometric analysis with a deuterium-labeled internal standard is without any doubt the method of choice.

See also: **Analytical Techniques**: Mass Spectrometry. **Drugs of Abuse**: Classification, including Commercial Drugs.

Further Reading

Braithwaite RA, Jarvie DR, Minty PSB, Simpson D and Widdop B (1995) Screening for drugs of abuse. I: Opiates, amphetamines and cocaine. *Annals of Clinical Biochemistry* **32**:123–153.

Cone EJ and Deyl Z (eds) (1992) Toxicological and forensic applications of chromatography. *Journal of Chromatography* **580**:1–325.

Department of Health and Human Services, Substance Abuse and Mental Health Services Adminstration (1994) Mandatory guidelines for federal workplace drug testing programs. *Federal Register* **59**:29908–29931.

Flanagan RJ, Braithwaite RA, Brown SS, Widdop B and de Wolff FA (1995) *Basic Analytical Toxicology*, pp. 1–273. Geneva: World Health Organization.

Goldberger BA and Cone EJ (1994) Confirmatory tests for drugs in the workplace by gas chromatography mass spectrometry. *Journal of Chromatography A* **674**:73–86.

Karch SB (ed) (1998) *Drug Abuse Handbook*. Boca Raton, FL: CRC Press.
Killander J and the Toxicology Experts' Working Group (1997) Recommendations for the reliable detection of illicit drugs in urine, with special attention to the workplace, in the European Union (December 1996). *Scandinavian Journal of Clinical and Laboratory Investigation* **57**:97–104.
Levine B and Smith ML (1990) Stability of drugs of abuse in biological fluids. *Forensic Science Review* **2**:147–157.
Liu RH and Gadzala DE (1997) *Handbook of Drug Analysis: Applications in Forensic and Clinical Laboratories*. Washington, DC: American Chemical Society.
Liu RH and Goldberger BA (eds) (1995) *Handbook of Workplace Drug Testing*. Washington, DC: AACC Press.
Maurer HH (ed.) (1998) Chromatography and capillary electrophoresis in forensic and clinical toxicology. *Journal of Chromatography B* **713**:1–279.
Simpson D, Braithwaite RA, Jarvie DR *et al.* (1997) Screening for drugs of abuse. II: Cannabinoids, lysergic acid diethylamide, buprenorphine, methadone, barbiturates, benzodiazepines and other drugs. *Annals of Clinical Biochemistry* **34**:460–510.
Tagliaro F and Deyl Z (eds) (1999) Toxicological analysis. *Journal of Chromatography B* **733**:1–280.
Wu AHB (1998) Integrity of urine specimens for toxicologic analysis: adulteration, mechanisms of action, and laboratory detection. *Forensic Science Review* **10**:47–65.

Antemortem Blood

M D Osselton, Forensic Science Service, Reading, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0488

Introduction

Blood is an exceptionally important tissue to the forensic toxicologist, as, in many cases, the concentrations of drugs in antemortem blood samples are interpretable with respect to the effect(s) that the drug may have been exerting on a person at the time of an offense. This article provides an outline of the nature of blood and the use to which it may be put by the toxicologist. The article also offers some thoughts on the considerations that must be taken into account before any meaningful interpretation can be made.

Sample Collection

Antemortem blood samples that are collected and submitted for forensic toxicological analysis usually fall within one of three categories:

1. The administration of a drug or poison to a party with the intent of willfully causing harm (e.g. attempting poisoning).
2. The administration of a drug to a party with the intent of producing incapacitation in them (e.g. sexual assault/date rape and doping).
3. To determine whether the donor of the sample may have been under the influence of a drug at the time an offence was committed (e.g. drunk driving or where a person claims not to have been responsible for his or her actions as a result of being under the influence of a drug).

The presence of a parent or unchanged drug in blood is, in general, often regarded as significant on the basis of the philosophy that if an unchanged drug is circulating in blood, it is potentially available to bind to, and subsequently exert some effect at, its site of action. This is of particular relevance in forensic toxicology where the interest lies with drugs that can influence behavior or exert a harmful or a life-threatening effect.

Blood is a particularly important tissue for the toxicologist because it provides the means whereby drugs, their metabolites and other toxic substances are transported from their point of entry into the body to the site(s) at which they may exert their various effects. Likewise, blood provides the transportation mechanism that assists with the removal of drugs and other substances from the various tissues in the body to their site(s) of elimination.

Since blood provides the transport mechanism that enables drugs and their metabolites to move around the body, the detection of a drug and/or its metabolites in blood may provide the toxicologist with information than can be interpreted with respect to the potential effects that may be exerted on the individual. In every case there are a number of things to consider before attempting to interpret analytical results obtained from an antemortem blood sample:

- the blood sample itself (e.g. whole blood/plasma, preservative);
- type of container used for sample collection and storage;
- analytical method used for detection/quantification;
- time of blood sampling in relation to the incident of interest;
- use of databases;

- pharmacology of the drug(s) – mode of entry into the body;
- actions on the body – drug interactions;
- case history;
- demographic factors;
- disease-related factors;
- extracorporeal factors.

Only after these basic issues have been considered should any attempt be made to interpret analytical results. One thing is certain: in a forensic situation where a living subject is concerned, you must get it right the first time. The sample will rarely be perfect or ideal but, unlike in clinical toxicology or therapeutic drug monitoring, where the patient can be requested to provide a second sample, it is not possible to go back for more at a later time or date. A blood sample that is collected in a forensic investigation is unique to the specific situation/incident that is under investigation.

The Blood Sample

Blood is an exceptionally complex tissue, comprising of red corpuscles (erythrocytes), white cells (leukocytes), proteins and other chemicals suspended in a fluid known as serum. It has been estimated that blood circulates completely around the body in approximately 2–3 min. The total blood volume of the body lies between 52–83 ml kg^{-1} body weight for an adult male and 50–75 ml kg^{-1} for an adult female or 2.5–4 l m^{-2} body surface area. The total plasma volume lies between 49 and 59 ml kg^{-1} body weight or 1.4 and 2.5 l m^{-2} body surface area. The interstitial fluid volume is approximately three times as great as the plasma volume.

Leukocytes are concerned principally with defense of the body against disease and infection and are not usually of direct interest to the forensic toxicologist, unless they have been measured clinically for the purpose of diagnosis of a patient's illness. Poisoning by proteinaceous toxins such as ricin may give rise to abnormally high white cell counts and, unless the poisoned victim is being monitored, the toxicologist would normally be unaware of the white cell count. The platelets are primarily involved with the clotting mechanism of blood and are not normally considered by the forensic toxicologist.

Erythrocytes are principally associated with the transportation of oxygen to the tissues and carbon dioxide from the tissues to the lungs. Erythrocytes are approximately 7.5 μm in diameter and contain no nucleus. The erythrocyte membrane is composed of approximately 40% phospholipids, cholesterol and glycolipids; 50% proteins; and 10% carbohydrates. Erythrocytes provide the red color to the blood, due to the presence of the complex molecule, hemoglobin, that they contain. Hemoglobin acts as the carrier molecule for gases such as oxygen, carbon dioxide and carbon monoxide. The complex nature of the erythrocytes makes them toxicologically interesting because interactions between toxins and hemoglobin or the red cell membrane can be very significant. Of the plasma proteins, the albumins play an important role as carriers of many drugs around the body.

Blood contains approximately 85% water (w/v) and has a specific gravity of 1.055. The percentage of the volume of the blood that is occupied by the cells is known as the hematocrit, that occupied by plasma is known as the plasmacrit. In males, the normal range of hematocrit values is 42–52%; females 37–47%; children 30–43%; and newborn infants 53–65%. Consideration of the hematocrit value may be significant in relation to drugs that bind to the erythrocytes, as subjects having abnormally low hematocrit values would have a low number of cells and subsequently a low hemoglobin concentration (i.e. have a lower drug-binding capacity).

The most frequently encountered blood samples in forensic toxicology are whole blood samples, largely because of the extended time intervals that elapse between sample collection and arrival at the laboratory. Clinical samples normally reach the laboratory within a few hours, at most, from the time of collection, permitting them to be separated before the blood begins to hemolyze. Forensic toxicology blood samples are usually hemolyzed on arrival at the laboratory, having spent some time in a police station refrigerator before being sent through the postal system and arriving at the laboratory some days after they were collected from the donor. Venous blood is the specimen of choice for forensic antemortem samples: it can normally be collected easily through a syringe needle or butterfly cannula from the forearm (cubital fossa) in a reasonable volume without causing too much discomfort to the donor.

There is a valid argument that, because the forensic toxicologist usually has no idea what drug or poison might be present in the sample, whole blood is the most appropriate sample to analyze. The analysis of whole blood facilitates the detection of drugs/poisons that may have associated with a particular blood component or fraction, whereas analysis of plasma only might lead to the analyst missing a substance that has preference for binding to erythrocytes. The analysis of whole blood is, however, more complex than the analysis of plasma and may require incorporation of a greater number of cleanup steps into the analytical process. Examples of drugs/poisons that have an affinity for erythrocytes are shown in **Table 1**.

Table 1 Drugs/poisons showing an affinity for erythrocytes

Atropine
Carbamazepine
Carbon monoxide
Chloroquine
Cyanide
Maprotiline
Pethidine in young persons
Salicylic acid
Meprobamate
Mercury
Propranolol

Sample Collection and Storage

The types of containers used for the collection of blood samples are of exceptional importance and can affect the efficiency of analysis and subsequent interpretation of analytical results. It is not always possible for the toxicologist to control the containers used for the submission of samples and all too often the doctor collecting the sample will use whatever sample container is at hand. Many laboratories issue kits for different types of cases, together with detailed instructions about what specimens to collect and how to transport them to the laboratory. Ideally, kits issued for the collection of antemortem blood samples should contain:

- a 10–20 ml syringe, with needle, of large enough volume to collect sufficient sample in one attempt;
- two 5 ml sealed glass septum vials with insert rubber septa containing a fluoride preservative at a concentration of $>2\%$;
- at least one 10 ml vial with a foil-lined cap containing fluoride oxalate preservative.

While clinical specimen vials containing lithium heparin or sodium oxalate do not appear to exert a significant effect on the analysis of common drugs of abuse, clinical tubes containing fluoride and designed for the purpose of collecting blood for glucose measurement are totally unsuitable for the storage of blood requiring alcohol analysis. Not only are the seals/stoppers not sufficiently secure to insure that no loss of the volatile alcohol occurs but the fluoride concentration is not sufficiently high to inhibit microbial activity. Since the toxicologist may occasionally be required to analyze lithium in an antemortem blood specimen, the use of either sodium oxalate, fluoride or tubes containing no preservative are more appropriate for general toxicological analysis than those containing lithium heparin. Caution must also be exercised when selecting swabs for cleaning the skin before blood sampling. Swabs containing any type of alcohol should be avoided lest they interfere with laboratory analysis for volatile substances.

Police surgeons should be encouraged to fill the septum vials so that as little space as possible is left between the specimen and the septum. It may be necessary to use a venting needle in order to achieve this, but many police surgeons are reluctant to use needles because of the risk of needle-stick injuries. However, this is the only satisfactory way to insure that volatile substances, such as alcohol or organic solvents, are not lost. Any remaining blood should be transferred to the screw cap vials. Glass containers are preferred to plastic containers because they offer greater resistance than plastic containers to the loss of volatile substances. Glass containers are also less likely to contain lubricants and plasticizers, which have been known to compete with and displace protein-bound drugs from their binding sites or interfere with analysis. Under no circumstances should plastic containers be used for collecting samples that may require analysis for volatile substances, such as toluene, as significant losses may occur owing to the permeability of plastic to the volatile substance. Care must also be taken not to store blood samples that have been submitted in plastic containers in an environment where contamination by environmental solvents might occur as a result of solvents diffusing into the blood through the walls of the tube.

Blood samples should ideally be stored refrigerated below 4°C while waiting to be analyzed. Blood samples should be analyzed as soon as possible after collection. The time between screening analysis and confirmatory/quantitative analysis should also be kept as short as possible. Knowledge about the stability of most drugs in blood is incomplete. Cocaine and benzoylecgonine are very unstable when stored for even short periods of less than 3 months. Cannabinoid concentrations are also known to fall off significantly, even when stored frozen. One study showed that the stability of morphine fluctuated over a 5 year period in blood samples stored at ambient temperature. The same study demonstrated that methamphetamine, amphetamine and phencyclidine (PCP) could all be detected in antemortem blood samples stored at ambient temperature for 5 years, although, on average, the concentrations of each of these drugs decreased over the period of the study.

Analytical Methodology

The German toxicologist, Manfred Moëller, once stated that 'it would be easier to get forensic toxicologists to use the same toothbrush than for them to use the same analytical method.' Judging by the numerous analytical methods available for the analysis of

drugs in whole blood, one could be forgiven for coming to this conclusion. Almost all laboratories tend to make modifications to published methodology to suit their own purposes. Forensic toxicologists, however, have a requirement placed on them by the courts to be able to interpret their results. In order to do so they must have confidence in the results and be able to demonstrate that if results from different laboratories are used for comparison purposes, they are truly comparable. All analytical methods should be fully characterized so that the analyst knows the specificity, accuracy, precision, limit of detection, limit of quantification, linearity, robustness, reproducibility and the range of concentrations over which the assay performance is valid.

Analytical methodology for forensic purposes is normally a two-stage process comprising an initial screening test, to determine whether any substance of interest or significance is present, followed by confirmatory, and possibly quantitative, analysis. The screening process used to determine whether common drugs of abuse may be present usually employs an immunoassay (subject to availability). Immunoassays used for the analysis of whole blood involve either radioisotopes (radioimmunoassay) or microtiter plates, where the drug antibodies are bound to the walls of the microcells at the point of manufacture. For most common drugs of misuse, immunoassays offering a rapid and relatively inexpensive screening stage, are available.

Although very sensitive, radioimmunoassay necessitates the laying aside of special laboratory areas for the storage and handling of radioisotopes and presents potential problems associated with the disposal of radioactive waste. Microplate assays offer a convenient and cost-effective means of screening whole blood samples that is readily automated. The microplate assay utilizes immobilized antibodies bound to the walls of the microplate wells to provide binding sites for drugs present in the sample. Drug substances present in the blood sample bind to the immobilized antibody during an initial incubation period. At the end of the incubation period, the blood sample is removed, the cells are washed to remove traces of blood and a further reagent is added; this develops a color reaction that depends upon the presence or absence of substances that have bound to the antibodies. The color intensity can then be read spectrophotometrically.

Drugs for which no immunoassays are available may be screened by gas chromatography (GC), high-performance liquid chromatography (HPLC) or capillary zone electrophoresis (CZE) after a preliminary extraction from blood. The advent of 'bench-top' mass spectrometry (MS) has made it possible for laboratories to combine the separating powers of GC or HPLC with the unique and sensitive identification properties of MS to enable screening and confirmation to be combined into a single drug identification procedure. The combination of a chromatographic separations technique and the ability of MS to yield a molecular fingerprint of most drugs is regarded as providing a high degree of certainty for the positive identification of drugs and/or their metabolites in blood. GC-MS and HPLC-MS also provide an opportunity for quantitative analysis to be undertaken.

Prior to confirmatory analysis by GC-MS or HPLC-MS, the drugs of interest must be isolated from the biological matrix of blood. The complex nature of whole blood compared with plasma means that many methods developed for clinical analysis and therapeutic drug monitoring are unsuitable for forensic analysis. It is not possible in this short monograph to detail the numerous extraction procedures published for purification of drugs/drug metabolites and the reader is recommended to refer to the specialist toxicological literature.

Interpretation

Interpretation is the ultimate objective of forensic toxicology analyses. Unless some interpretation can be applied to analytical results, there is little point in undertaking analysis. The interpretative process is the most interesting and challenging aspect of forensic toxicology, and that which is most open to challenge and discussion. The most frequently asked questions of the forensic toxicologist are:

- Was a drug/poison present or absent?
- What sort of action could the substance detected have exerted on the donor of the sample?
- If a drug/poison was detected, was it present in sufficient quantity to affect the behavior or well-being of the donor of the sample?
- Could the substance detected have been influencing the donor of the sample at the time of an alleged incident?

Qualitative analysis will usually be sufficient to allow a statement about whether a drug was present above the limit of detection of the analytical method; however, before embarking on expensive quantitative analysis, the toxicologist must consider whether anything meaningful can be gained by it. Drugs affect different people in different ways and can even affect the same person in different ways on different occasions. The concentrations of a number of common drugs in blood and plasma are summarized in **Table 2**, at what are normally regarded as therapeutic concen-

Table 2 Therapeutic and toxic concentrations of some common drugs of forensic interest in blood and plasma.

Drug	*Percentage in plasma*	*Therapeutic plasma ($\mu g\,ml^{-1}$)*	*Therapeutic whole blood ($\mu g\,ml^{-1}$)*	*Toxic plasma (P) Whole blood (B) ($\mu g\,ml^{-1}$)*
Acebutolol		0.5–1.25		
Amitriptyline	65	0.04–0.4		>4 (B)
Alprazolam		0.02–0.06		0.06–0.4 (P)
Amphetamine	50	0.03–0.5		0.5–1 (P)
Amylobarbitone		0.1–2.2		>2.3
Arsenic	50		<0.05	>0.1
Atenolol		0.2–0.6		>2
Atropine		0.002–0.025		0.02–1.0
Barbitone		15–40		60–100
Butobarbitone		5–15		20–80
Caffeine		2–9		20–30
Carbamazepine		5–10		12–44
Carbon monoxide		<7% sat		>25%
Chlormethiazole	41	2		10
Chloroform		1–5 (20–232 in anesthesia)		>70
Chlorpheniramine		0.003–0.02		>20
Chloroquine		0.01–0.4		>0.5
Chlorpromazine		0.05–0.5		1–35
Clomipramine	62	0.24–0.7		
Cimetidine		0.5–1		
Clonazepam		0.013–0.072		>0.1
Cocaine (BE)			0.03–0.3	
Codeine	80		0.03–0.1	1.3–5
Cyanide			0.02–0.4	>0.2
Cyclizine		0.03–0.3		0.75
Desipramine		0.01–0.28		0.5–1.56
Dextropropoxyphene			0.05–0.4	>0.8
Desmethyldiazepam	89	0.1–1.7		5
Diazepam	98	0.1–1.0	0.5–2.5	5–30
Digitoxin		0.003–0.04		>0.035
Digoxin		0.001–0.002		>0.003
Dihydrocodeine		0.024–0.13		>0.8
Dipipanone			0.17	
Dothiepin		0.05–0.2		
Doxepin		0.03–0.15		
Flunitrazepam		0.005–0.015		0.05
Flurazepam (as desalkly metabolite)		0.01–0.14		
Glutethimide		2.9–12.2		12–20
Gamma hydroxybutyrate		24–158		>156
Imipramine	53	0.04–0.105		>0.4
Lignocaine		2–5		7–14
Lithium*		0.03–1.4*		>1.4*
Lorazepam		0.017–0.07		
Lormetazepam		0.001–0.018		
Medazepam			0.01–0.17	0.6
Methadone			0.03–0.34	0.4
Methamphetamine		0.01–0.05		0.2–1
Methyl thioamphetamine (MTA)		0.13–0.76		
Morphine		0.08–0.1		0.1–0.5
Nortriptyline		0.05–0.25		0.2
Paracetamol		10–20	10–25	30–100
Paraldehyde		30–150	30–300	200–400
Pentazocine	50		0.03–0.16	>0.2
Pethidine		0.1–0.8	1.5	
Phenobarbitone	50	10–40	6–40	40–60
Propranolol		0.05–0.1		>2
Quinalbarbitone		1.8–5.3		4–16
Quinidine		2–5		>10
Quinine	50	1.7–1.8		
Salicylic acid		20–250		210–280

Table 2 continued

Drug	*Percentage in plasma*	*Therapeutic plasma* ($\mu g\,ml^{-1}$)	*Therapeutic whole blood* ($\mu g\,ml^{-1}$)	*Toxic plasma (P) Whole blood (B)* ($\mu g\,ml^{-1}$)
Temazepam	100	0.6–0.9		>1.0
Theophylline	66	11–21		>20
Thiopentone		1–5		>10
Thioridazine		0.2–1		>2
Triazolam		0.003–0.02		
Trimipramine		0.07–0.3		>0.5
Tubocurarine	100	0.04–6.0		
Warfarin	98	1.2–7		10–12
Zopiclone		0.01–0.05		10–12
BE, benzoylecgonine				

* mmol/litre

trations and at concentrations where some toxic effect may be experienced. Since, however, many commonly misused drugs are not, or have not been, used clinically and have no therapeutic use, there are no 'therapeutic' concentrations available that can be substantiated by clinical trials. Most drugs of misuse are taken because they exert a psychotropic effect (i.e. they influence or alter the way a person thinks, behaves, perceives his or her surroundings or reacts to external stimuli). It is these drugs that the forensic toxicologist is requested to analyze and comment upon most frequently.

As most of the therapeutic/toxic concentrations of drugs in blood have been measured and published by clinical laboratories, the data in Table 2 relates primarily to plasma drug concentrations rather than whole blood. Information concerning the distribution of drugs between plasma and whole blood is also very limited. Hence, analysts using the data in Table 2 should exercise caution in trying to extrapolate between plasma and whole blood concentrations. Where data have been made available concerning whole blood and plasma drug concentrations, this has been incorporated in the table. It must be noted that the ranges of results listed under 'therapeutic' and 'toxic' should be seen as approximate because they have been collated from papers published in the scientific literature where authors will have used different methods of analysis, and there is no definable point below which a drug exerts a therapeutic effect and above which it is toxic.

Information

The extent to which any reasonable attempt at interpretation can be made depends on the information available to the toxicologist, who should always insist on being provided with as much information as possible before starting a case and attempting to make any interpretation. A number of important questions to ask submitting officers are listed in **Table 3**. The toxicologist should also strive to keep as up to date as possible with current scientific literature so that the results and case history may be related to the known pharmacology and pharmacokinetics of the drug in question. In particular it is important to consider:

Table 3 Minimum information that should be requested to accompany antemortem blood samples requiring toxicological analysis.

1. Name
2. Date of birth
3. Race (e.g. Caucasian / Chinese / Negroid / Other – specify)
4. Sex
5. Body weight
6. Height
7. Build: slim / proportionate / stocky and muscular / stocky and obese / obese
8. Full description of circumstances leading up to arrest
9. Behavior / behavioral characteristics, i.e. speech, stance, appearance (tidy / dishevelled), eye pupils (wide or narrow), shaking, tremors, shivering, face color (ruddy or ashen)
10. Comments of police surgeon
11. Recent medical history – any heart trouble or diabetes?
12. Medications taken within preceding 10 days:

    prescription
    over-the-counter medications/preparations
    cold / flu remedies
    illegal/illicit drugs

13. Any consumption of alcohol: record all drinks and times, using a 24 h clock, and drinking pattern (obtain evidence in form of cans/glasses if possible)
14. Results of breath alcohol tests if available (copies of printouts should be submitted)
15. Time and description of last meal
16. Occupation: availability / exposure to any occupational toxins or solvents

- mode of entry of the drug into the body;
- distribution of drug within the body;
- metabolism of drug within the body;
- excretion profile of the drug;
- actions exerted by the drug on the body;
- potential drug interactions;
- whether tolerance may be induced to the effects of the drug and, if so, how this might affect the individual and the subsequent interpretation of results.

Medicinal drugs are normally administered either orally, transdermally or by injection (intravenous or intramuscular). The route by which a drug enters the body can significantly influence the time of onset and intensity of its effect. Since psychotropic drugs exert their psychotropic effects by acting on the central nervous system, the most intense and desirable effects are obtained by delivering the active drug in an optimal concentration to the brain as quickly as possible. As cocaine, tetrahydrocannabinol and heroin/morphine undergo extensive first-pass metabolism by the liver when taken orally, they are usually introduced into the body by injection, smoking (heroin, crack cocaine, cannabis) or nasal insufflation (cocaine).

Where the elimination half-life is known, it is sometimes possible to extrapolate backwards from the concentration of drug measured in an antemortem blood sample collected at a known time and to estimate what the concentration of drug might have been at the time of an incident. Any such calculations are based on a number of assumptions, and before any attempt to use such calculations is undertaken, a fully authenticated case history should have been obtained and all appropriate caveats should be included in the statement of interpretation.

Unless robust information is available to justify any interpretation based on a pharmacokinetic back-calculation, the toxicologist should avoid being led into undertaking such a calculation by investigating officers or lawyers. The toxicologist who offers an interpretation that cannot be fully justified is laying him- or herself open for an uncomfortable and potentially discrediting experience under cross-examination in court.

See also: **Serology**: Blood Identification. **Crime-scene Investigation and Examination**: Preservation.

Further Reading

Baselt RC (1997) *Disposition of Toxic Drugs and Chemicals in Man*, 5th edn. Foster City, CA: Chemical Toxicology Institute

Buris CA and Ashwood ER (1986) *Tietz Textbook of Clinical Chemistry*. Philadelphia: WB Saunders.

DeZeeuw R, Muller BS and Franke JP et al (1994) *Laboratory guidelines for toxicological analysis. Contributions to Forensic Toxicology*. Proceedings of the 31st International Association of Forensic Toxicologists, 1993, Leipzig. Leipzig: Molin.

Dyer JE (1991) γ-Hydroxybutyrate: a health-food producing coma and seizure-like activity. *American Journal of Emergency Medicine* **9**:321–324.

Giorgi S and Meeker J (1995) A five year stability study of common illicit drugs in blood. *Journal of Analytical Toxicology* **19**:392–398.

Osselton MD (1985) Toxicological tables – a compendium of pharmacological, therapeutic and toxicological data on 137 drugs and chemicals in man. In: Sunshine I (ed.) *Methodology for Analytical Toxicology*, vol. 3. Boca raton, FL: CRC Press.

Osselton MD (1991) *Good laboratory practice*. Proceedings of the 29th International Meeting of the International Association of Forensic Toxicologists Copenhagen, pp. 93–94.

Repetto MR and Repetto M (1997) Therapeutic, toxic and lethal concentrations in human fluids of 90 drugs affecting the cardiovascular and hematopoietic systems. *Clinical Toxicology* **35**(4):345–351.

Body Fluids

S Kerrigan, California Department of Justice, Sacramento, CA, USA
B A Goldberger, University of Florida College of Medicine, Gainesville, FL, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0487

Introduction

The analysis of drugs of abuse in unconventional biofluids is a rapidly expanding area of toxicological science. Blood and urine have long been the dominant matrices for drugs of abuse detection. Incorporation of drugs into these biofluids is well understood and the analysis and interpretation of these findings has become routine for most commonly abused drugs. Recently, however, interest has shifted towards alternative specimens that may offer distinct advantages over conventional biofluids. These benefits may include long-term, cumulative information on drug use, or the convenience of noninvasive sample collection.

Following absorption of a drug, distribution, metabolism and excretion pathways account for the

sequential appearance of drug and metabolite in different tissues and fluids. Physicochemical characteristics of the drug and biofluid can be used to rationalize or predict the appearance of drug in a particular biofluid or compartment of the body. The pK_a, lipid solubility, protein binding and biofluid composition determine the extent to which the drug is present. Transfer of drug from the circulating blood plasma (pH 7.4) to another biofluid involves transport across membranes, which are effective barriers against ionized, highly polar compounds. Following penetration of the membrane and transfer into the biofluid, the pH differential may result in ionization of the drug, restricting further mobility. Accumulation of the drug in this way is commonly referred to as 'ion trapping'.

The presence of drug in a biofluid indicates exposure to the substance, perhaps unwittingly. Drugs of abuse are generally self-administered by oral, intranasal, intravenous or smoked routes. Passive smoke inhalation and ingestion of certain foodstuffs, such as poppy seeds or hemp oil, may result in detectable amounts of drug in certain biofluids. Unconventional drug exposure, such as that which occurs *in utero* or in nursing mothers, has also necessitated the use of alternative biological specimens. Our discussion of these unconventional samples is limited to amniotic fluid, breast milk, saliva, semen and sweat. The growing concern regarding the effects of drugs on health and human performance has highlighted alternative biofluid analysis in multiple forums: law enforcement, probation, parole, drug compliance and abstinence programs, employment, health and insurance, among others. Combined interest and growing expertise in alternative biofluid analysis has increased momentum in the field. The relative expectations, limitations and interpretation vary widely between these different applications. Recent advances in the analysis of alternative biofluids and matrices have accelerated drugs of abuse detection in some of these areas. Of critical importance is the choice of biofluid for analysis, as each may provide unique chemical and pharmacologic information. There are a great many factors which may influence the choice of biofluid for drug analysis and some of these are listed in **Table 1**. The interpretive value, advantages and disadvantages of each biofluid are summarized in **Table 2**.

Table 1 Factors influencing the choice of a biofluid

Sample collection
Invasiveness
Risk of infection, complication and hazards
Protection of privacy
Ease and speed of collection
Training of personnel (medical/nonmedical)
Likelihood of adulteration
Contamination
Volume of specimen
Analysis
Qualitative or quantitative
Window of detection
Drug concentration/accumulation in biofluid
Parent drug or metabolite(s)
Stability of drug analytes
Biofluid storage requirements
Pretreatment of specimen
Limitations of the matrix
Likelihood of interferences
Inter-and intrasubject variability of the matrix
Use of existing analytical procedures
Speed of analysis
Personnel training requirements
Appropriate cut-off concentrations
Interpretation
Pharmacologic effects
Behavioral effects
Indicator of recent drug use (hours)
Short-term drug exposure (days)
Long-term drug exposure (weeks)
Forensic defensibility

Biofluids

Amniotic fluid

Increased use of illegal drugs by expectant mothers has led to the need for prenatal toxicological testing. Exposure to drugs of abuse has demonstrated effects on both the fetus and the neonate. Higher rates of fetal distress, demise, growth retardation and adverse neurodevelopment have been documented. Cocaine, heroin, amphetamines, and nicotine have been associated with impaired fetal growth and acute withdrawal syndromes. The greatest risk of neonatal abstinence syndrome occurs with narcotic drugs but has also been observed with cocaine and amphetamines. In a 1997 study conducted in Michigan, in which gestational drug exposure was measured in nearly 3000 newborns, as many as 44% of neonates tested positive for drugs. Of these, 30.5% tested positive for cocaine, 20.2% for opiates and 11.4% for cannabinoids.

Anatomy/physiology Amniotic fluid, which is produced by cells that line the innermost membrane of the amniotic sac (amnion), is the liquid that surrounds and protects the embryo during pregnancy. This fluid cushions the fetus against pressure from internal organs and from the movements of the mother. Production of fluid commences the first week after conception and increases steadily until the 10th week, after which the volume of fluid rapidly increases.

Table 2 Advantages and disadvantages of biofluids

Biofluid	*Advantages*	*Disadvantages*
Amniotic fluid	• Minimal sample preparation • Amenable to most analytical techniques • Relatively few interferences • Useful in determining *in utero* exposure at an early stage of development (16–20 weeks)	• Highly invasive medical sampling procedure • Requires local anesthetic, ultrasound scan and highly trained personnel • Risk of complication associated with sampling
Breast milk	• Many drugs present • Maternal and neonatal drug exposure can be assessed	• High lipid content may interfere with analysis • Additional extraction steps may be required • Disposition of drug varies with milk composition • Matrix variability between individuals and in one feed • Inconvenient sample collection • Invasion of privacy
Saliva	• Noninvasive sample collection • Easily accessible, speedy collection, ideal for on-site tests • Minimal sample preparation • Few interferences • Amenable to most analytical techniques • Many drugs determined • Parent drug present • Correlation with free drug concentration in blood • Pharmacologic interpretation possible • Indicates recent drug use	• Low concentration of drug relative to urine • Drug concentration varies with saliva flow rate and pH • Sample collection technique may influence drug disposition • Small sample volume (1–5 ml) • Short drug detection time (similar to blood) • Unsuitable matrix for determining drug history • Possibility of sample adulteration by mouth • Contamination of saliva after smoking or orally ingested drug • Drug instability/transformation
Semen	• Minimal sample preparation • Amenable to most analytical techniques	• Invasion of privacy • Heterogeneous composition of fluid (intra- and intersubject variation)
Sweat	• Noninvasive specimen collection • Easily accessible biofluid • Rapid specimen collection (skin wipe) • Convenient collection device (patch) • Patch may be worn for up to 14 days • Many drugs determined • Parent drug present • Provides a cumulative estimate of drug exposure • Useful in determining individual drug history • Less frequent drug testing compared to urine • Adulteration unlikely	• Small sample volume • Volume of sample may be unknown • Requires extraction prior to analysis • Skin irritation or discomfort possible • Patches must be worn for extended period (days) • Drug stability/degradation in sweat patch • Low concentration of drug • Generally short window of detection in skin wipes • Drug disposition varies with flow rate and pH • Volume excreted varies between individuals, emotional state, temperature, exercise • Unequal distribution of sweat glands • Drug concentration unlikely to correlate with blood • Pharmacologic interpretation unlikely • No correlation with impairment

Amniotic fluid, which may total about 1 liter at 9 months, contains cells and fat that may give the liquid a slightly cloudy appearance. Maternal drug use may expose the fetus to the drugs and their metabolites, the toxicity and effects of which may be detrimental to the normal development of the fetus. Amniotic fluid is constantly circulated, being swallowed by the fetus, processed, absorbed and excreted by the fetal kidneys as urine at rates as high as $50\,ml\,h^{-1}$. This circulation of fluid continuously exposes the fetus to compounds that may absorb in the gut or diffuse through fetal skin in the early stage of development. The encapsulation of the fetus in this fluid may prolong exposure to harmful drugs or

metabolites. Reduced metabolic or enzyme activity in pregnant women or the developing fetus may decrease drug metabolism, and therefore, compound the risks. The pharmacokinetics of drug disposition *in utero* varies from drug to drug, and the acute and chronic effects that may result are the topic of continuing research. In addition, indirect effects of maternal drug use, such as those on uterine activity and maternal circulation, may be as damaging as the drug itself.

Sampling The collection of amniotic fluid (amniocentesis) usually takes place between the 16th and 20th weeks of pregnancy. The liquid is usually collected to test for fetal abnormalities. The presence of illicit drugs or their metabolites in amniotic fluid suggests that the fetus has been exposed to these substances via maternal blood circulation. A maternal serum sample taken at the same time as the test may provide complementary toxicological data and help assess the relative risk to the fetus. Of the alternative biofluids described here, amniocentesis is perhaps the most invasive specimen collection procedure. Prior to amniocentesis, an ultrasound scan is used to determine the position of the fetus. After receiving a local anesthetic, a needle is inserted through the abdomen into the womb where there is the least chance of touching the placenta or the fetus. Although complications are rare, miscarriage occurs in approximately 1% of women. Typically 5–30 ml of amniotic fluid is removed; it is slightly alkaline in nature. However, the pH of amniotic fluid decreases during pregnancy due to fetal urination, reaching near neutral pH at full term. Amniotic fluid, which is 99% water, contains dilute plasma components, cells and lipids.

Analysis Drugs present in amniotic fluid can be analyzed using well-established techniques that are routinely used for blood and urine. A variety of analytical methods that have been employed are listed in **Table 3**. Drugs of abuse have been detected in human amniotic fluid using screening tests, such as immunoassays, and confirmatory methods such as gas chromatography–mass spectrometry (GC-MS). Few interferences are encountered with amniotic fluid due to its high water content. Sample pretreatment may not be necessary before the immunoassay screening test. However, confirmation of presumptive positive results by GC-MS requires isolation of the drug by liquid–liquid or solid-phase extraction.

Interesting/relevant findings A number of maternal, fetal and placental factors may affect fetal drug exposure. Of these, binding to serum proteins in the maternal and fetal circulation and fetal elimination are particularly important. Accumulation of unmetabolized drugs in amniotic fluid occurs as a result of fetal renal excretion, and perhaps via the fetal membranes. However, elimination of the drug by the fetal liver may be less important than in the adult. Polar conjugated drug metabolites may accumulate in the fetus or amniotic fluid owing to limited placental transfer.

Table 3 Methods of analysis of drugs in body fluids

Purification
Liquid–liquid extraction
Solid-phase extraction
Supercritical fluid extraction
Drug detection by immunochemical techniques
Enzyme immunoassay
Enzyme-linked immunosorbent assay
Fluorescence polarization immunoassay
Radioimmunoassay
Drug identification by chromatographic techniques
Capillary electrophoresis
Gas chromatography
Gas chromatography/mass spectrometry
Gas chromatography/mass spectrometry/mass spectrometry
High-performance liquid chromatography
Liquid chromatography/mass spectrometry
Thin-layer chromatography

Small lipid-soluble drugs can rapidly diffuse across the placental barrier, producing similar drug concentrations in amniotic fluid and fetal plasma. Larger, water-soluble compounds, which are transferred more slowly, are incorporated into the amniotic fluid from fetal urine. Basic drugs may accumulate in the amnion due to ion trapping, resulting in drug concentrations in excess of those found in fetal or maternal plasma.

Maternal cocaine abuse is known to decrease uterine blood flow, cause fetal hypertension, cardiovascular effects and a deficiency of oxygen in the arterial blood. Retarded growth, congenital abnormalities, withdrawal syndrome, and cerebral hemorrhage or infarction have also been observed. Prenatal drug tests have shown amniotic fluid to contain cocaine and its major metabolite, benzoylecgonine, at concentrations up to 24 and 836 $ng\,ml^{-1}$, respectively, although concentrations as high as 3300 and 1600 $ng\,ml^{-1}$ have been measured following maternal death. The fetus is exposed to cocaine by maternal circulation as well as placental and fetal metabolism. After crossing the placental barrier by simple diffusion, the drug distributes between fetal and maternal blood. The amniotic sac and its contents serve as a deep compartment, with restricted, slow equilibrium between adjacent compartments. As a result, amniotic fluid inside this protective sac may expose the fetus to potentially harmful drugs or metabolites. Ion trapping and under-

developed renal function may result in an accumulation of the drug in fetal blood, thus compounding the risk.

Narcotic analgesics are reported to cross the placental barrier rapidly, but at physiological pH, when the drug is mostly charged, the concentration of the drug in the amniotic fluid is expected to be lower than that of the maternal plasma. Due to their high lipid solubility and lack of ionization, benzodiazepines readily cross the placenta. However, drug concentrations in the amniotic fluid remain low due to extensive protein binding in the maternal plasma and minimal renal excretion by the fetus. Drugs of abuse and their metabolites that have been detected in amniotic fluid and other matrices are summarized in **Table 4**.

Interpretation of toxicological findings is limited because drug dose, route, and time of administration are usually unknown. Long-term implications of prenatal drug exposure are limited and many consequences of fetal drug exposure are still unknown. Despite adequate understanding of the maternal consequences of drug abuse, fetal consequences for many drugs are poorly understood and this is a challenging area of maternal–fetal medicine.

Breast milk

The prevalence of drug abuse among pregnant women throughout urban America is reported to be between 0.4 and 27%, depending on geographical location. In the last two decades there has been nearly a threefold increase in the number of women who breast-feed their infants. These factors together significantly increase the likelihood of drug exposure that may result in acute toxicity or withdrawal syndrome. Although it is generally accepted that drug exposure of the infant from milk is less harmful than *in utero* exposure of the fetus, the overall effect on health and development is largely unknown. According to the American Academy of Pediatrics, amphetamine, cocaine, heroin, marijuana and phencyclidine (PCP) are considered unsafe for nursing mothers and their infants.

Anatomy/physiology The female breast consists of 15–20 lobes of milk-secreting glands embedded in the fatty tissue. During pregnancy, estrogen and progesterone, secreted in the ovary and placenta, cause the milk-producing glands to develop and become active. The ducts of these glands have their outlet in the nipple; by midpregnancy, the mammary glands are prepared for secretion. Colostrum, a creamy white to yellow premilk fluid, may be expressed from the nipples during the last trimester of pregnancy. This fluid, which is a rich source of protein, fat, carbohydrate and antibodies, is replaced with breast milk within 3 days of delivery of the fetus and placenta. Proteins, sugar and lipids in the milk provide initial nourishment for the newborn infant. The production of between 600 and 1000 ml of milk per day by the milk-secreting cells is stimulated by the pituitary hormone, prolactin.

Sampling/analysis Fluid is collected using a special device such as a breast-milk pump, after which established analytical techniques may be used to detect drugs of abuse. Breast milk, which contains protein (1%), lipid (4%), lactose (7%) and water (88%), varies in pH between 6.35 and 7.35. However, the high lipid content of milk may interfere or decrease the extraction efficiency or recovery of drug. Additional washing with nonpolar solvents such as hexane may be necessary to remove excess lipids before

Table 4 Partial listing of drugs and metabolites detected in biofluids

Biofluid	*Drug/drug metabolite*
Amniotic fluid	Benzodiazepines, benzoylecgonine, cocaine, cocaethylene, ecgonine methylester, meperidine, methadone, phencyclidine
Breast milk	Amphetamine, benzoylecgonine, clonazepam, cocaine, cocaethylene, codeine, chloral hydrate, diazepam, dothiepin, fentanyl, fluoxetine, 11-hydroxytetrahydrocannabinol, lithium, morphine, methadone, nordiazepam, oxycodone, phenobarbital, temazepam, Δ^9-tetrahydrocannabinol, 11-nor-9-carboxy-Δ^9-tetrahydrocannabinol
Saliva	Anhydroecgonine methylester, amphetamine, amobarbital, barbital, benzoylecgonine, buprenorphine, chloral hydrate, chlordiazepoxide, clonazepam, cocaine, codeine, *N*-desmethyldiazepam, diazepam, ecgonine methylester, ephedrine, heroin, hexobarbital, hydromorphone, 6-acetylmorphine, meperidine, methadone, methamphetamine, methaqualone, morphine, nitrazepam, phencyclidine, phenobarbital, phenylpropanolamine, phentermine, pholcodine, secobarbital, Δ^9-tetrahydrocannabinol, 11-nor-9-carboxy-Δ^9-tetrahydrocannabinol
Semen	Amphetamine, benzoylecgonine, cocaine, methadone, phenobarbital
Sweat	Amphetamine, benzoylecgonine, buprenorphine, cocaine, codeine, diazepam, dimethylamphetamine, ecgonine methylester, heroin, 6-acetyl morphine, methadone, methamphetamine, methylenedioxyamphetamine, methylenedioxyethylamphetamine, morphine, nordiazepam, oxazepam, phencyclidine, phenobarbital, Δ^9-tetrahydrocannabinol

chromatographic analyses. The effect of natural emulsifying agents in breast milk, which have detergent-like activity, may interfere with antibody–antigen reactions which take place in immunoassay screening tests. The daily variation of breast milk composition, combined with drug dose and time of administration relative to the expression of milk, is likely to affect the amount of drug present and the effect on the infant. Metabolic function and overall health of the nursing mother also play an important role. The concentration of drug found in the breast milk is subject to both within- and between-subject variation, further confounding attempts to generalize infant risk assessment. The lipid content of the milk varies not only daily, but also during a single feed; the latter portion of expressed milk may contain a severalfold increase in fat, which in turn may increase or decrease the concentration of a particular drug.

Interesting/relevant findings The transfer of drug into the milk depends on metabolism, protein binding and the circulation of blood in the mammary tissue. Passive diffusion transports the drug across the mammary epithelium into the milk. The mildly acidic pH of breast milk tends to trap weakly basic drugs, particularly those of low molecular weight, which can diffuse fairly rapidly through small pores in the semipermeable membrane. Although drugs that are extensively protein bound may not readily pass into the milk, emulsified fats serve to concentrate lipid-soluble drugs such as marijuana and PCP. For this reason, PCP has been detected in breast milk for as long as 41 days after cessation of maternal drug use. Not all lipophilic drugs produce deleterious effects in the nursing infant. Fentanyl, which was found to concentrate in lipid-rich colostrum at much higher concentration than maternal serum, has a low oral bioavailability, somewhat minimizing the risk to the child. However, inactive conjugated metabolites present in the mother's milk may undergo reactivation by deconjugation in the gastrointestinal tract of the infant.

Opiate addiction and withdrawal symptoms have been reported in infants receiving milk from substance-abusing mothers. Low oral doses of morphine to nursing mothers may produce drug concentrations in milk as high as $100\,\mathrm{ng\,ml^{-1}}$. Methadone maintenance ($50\,\mathrm{mg\,day^{-1}}$) of a drug-dependent nursing mother produced breast-milk concentrations between 20 and $120\,\mathrm{ng\,ml^{-1}}$ in the first 24 h after the dose, substantially lower than maternal plasma concentration. Benzodiazepines, which are excreted in the milk, can accumulate in the infant owing to underdeveloped metabolic and excretory function. Water-soluble drugs appear least likely to partition into the milk and are less likely to accumulate in the infant. However, stimulants such as amphetamine have been detected in milk at concentrations 3–7 times higher than those found in maternal plasma owing to ion trapping of basic drugs. Cocaine is also believed to preferentially partition into milk, increasing the likelihood of infant toxicity.

Despite the fact that some lactating women have reported using marijuana, little is known of the effect of this drug on the infant. The concentration of the active ingredient, Δ^9-tetrahydrocannabinol (THC), is reported to be as much as eightfold higher in breast milk than maternal plasma. Metabolites of THC and the parent drug were measured in breast milk at concentrations as high as $340\,\mathrm{ng\,ml^{-1}}$ in a chronic marijuana smoker.

Saliva

Saliva, which can be collected noninvasively, conveniently and without invasion of privacy, is perhaps one of the most appropriate biofluids for on-site drug testing. The active constituent of marijuana, which is the most widely abused drug in America, was detected in the saliva of 9% of motorists who displayed erratic driving behavior. Law enforcement agencies are particularly interested in saliva testing as a complementary tool for identifying impaired drivers with roadside drug testing devices. This approach, which is already in widespread use for employment, health and insurance drug testing, is rapidly gaining in popularity.

Anatomy/physiology Saliva is the clear viscous fluid that is secreted by salivary glands and mucous membranes that line the mouth. It serves to lubricate the oral cavity, assists in the swallowing of food and facilitates our sense of taste. This biofluid, which is comprised of 90% water, contains only about 0.3% protein. The remainder is made up of electrolytes, mucin, urea, lipids, digestive enzymes, white blood cells and debris from the lining of the mouth. Three pairs of glands secrete saliva via ducts in the mouth. Each gland consists of thousands of saliva-secreting sacs and a network of canals that transport the fluid into the main ducts in the mouth. The parotid glands, which are the largest of the salivary glands, lie inside the cheek, just below and in front of the ear. These glands secrete serous fluid that is derived from blood plasma. Serous fluid and mucin are secreted by the sublingual glands, located on the floor of the mouth beneath the tongue, and by the submandibular glands, which are just below the jaw on the front of the neck. Mixed saliva is comprised mostly of submandibular secretions (71%). Parotid and sublingual glands are responsible for 25% and 4% of the remaining

volume. This biofluid, which may be secreted at rates of 1.5 l day^{-1}, has an average pH of about 6.8.

Sampling Mixed saliva is collected noninvasively by expectoration, aspiration by vacuum, or by saturation of a cotton swab. Secretions from a specific gland may be collected using a special device or by cannulation, but this is uncommon. Salivation for the purpose of specimen collection may be increased by chewing or sucking an inert substance, such as PTFE tape, parafilm (wax) or a rubber band. It is necessary to ensure that no adsorption takes place between the drug and the chewed substance. Acidic candy or citric acid has also been used to stimulate glandular secretions. Care must be taken that residual food or drink in the mouth does not interfere with the analysis. It is possible that interfering substances could be placed in the mouth to adulterate the saliva, but there have been no known reports as yet.

Analysis Saliva contains few interferences and is amenable to most types of toxicological analyses. It is possible that endogenous enzymes in saliva could interfere with colorimetric or fluorescence immunoassays that detect drug indirectly from the activity of a labeled enzyme. If saliva is collected by expectoration or free flow, sample pretreatment may not be necessary before immunochemical testing. Dilution of the sample with buffer may be sufficient, although centrifugation may be used to remove any solid debris. Liquid–liquid or solid-phase extraction is necessary before chromatographic analysis.

Interesting/relevant findings The detection times for most drugs in saliva are fairly short relative to urine and do not provide a history or long-term profile of drug use. However, the parent drug may be detected for several hours following drug use at concentrations proportional to those measured in plasma. Saliva, which is an ultrafiltrate of interstitial fluid, contains the unbound fraction of drug, which is pharmacologically active. As a result, saliva tests may allow pharmacologic interpretation of results, based on the concentration of drug that was circulating at the time of the test.

Salivary ducts in the mouth are separated from the systemic circulation by a layer of epithelial cells. Transfer of drug from the plasma into saliva requires transport across the lipid membrane of these cells. Passive diffusion is perhaps the most important route of passage for most drugs, although ultrafiltration and active secretion from the blood may also occur. Transfer across cell membranes is restricted by size, ionization and macromolecular binding. Therefore the pH of the saliva, pK_a of the drug and plasma protein binding of the drug, strictly control the passage of drug into this biofluid. At fixed salivary pH (6.8), theoretical estimates of the partition of a drug between the saliva and plasma (S:P ratios) can be predicted from the Henderson–Hasselbach equation. Increasing the saliva flow rate can increase pH from 5.5 to 7.9, greatly affecting the disposition of drugs that are ionized at normal plasma pH. Stimulation of saliva flow by chewing waxed film or sour candy may decrease the concentration of weakly basic drugs, such as cocaine (pK_a 8.7), by severalfold. This weakens the reliability of saliva testing for pharmacologic interpretation and accounts for discrepancies in S:P ratios between authors. Although salivary pH is the principal determining factor for ionizable drugs, compounds, which are uncharged in plasma, are unaffected. In contrast to cocaine, the concentration of zwitterionic metabolite, benzoylecgonine (pK_a 2.3, 11.2), is unchanged by salivary pH.

Passive diffusion through lipid membranes requires the drug to be in a lipid-soluble form. Glucuronides or polar metabolites may be too hydrophilic to cross the membranes. Saliva has a minimal protein binding capacity compared with plasma. In order to be retained in saliva, the drug must be water-soluble, a property largely augmented by ionization of the drug. This transformation prevents back-diffusion of the drug into the plasma. Drugs of abuse that are strongly basic in nature may be preferentially distributed into saliva, as is the case with amphetamine. Salivary concentrations may exceed those measured in plasma by as much as two- to threefold. These types of drugs may be detectable for slightly longer periods of time, up to 2 days for amphetamine and methamphetamine.

The range of saliva pH is typically much narrower (6.5–7.2) than that of urine (4.5–8.0), suggesting that this alternative biofluid may be of diagnostic value for drugs whose excretion is heavily dependent on urinary pH, such as the amphetamines or PCP.

Numerous other drugs have been measured in saliva, including a number of opiates and synthetic opioids. Methadone, which is used to maintain opioid-dependent individuals, has been measured in saliva, where its concentration correlates well with that in plasma. Phenobarbital, amobarbital and other barbiturates have been detected in saliva, typically with S:P ratios of about 0.3. Other sedative drugs, methaqualone and meprobamate, have S:P ratios of 0.1 and 1, respectively. Owing to their widespread abuse, a number of benzodiazepines, including diazepam, have been investigated. S:P ratios were typically low (0.04 or less), which is very much lower than predicted. This is likely to be the result of the extensive plasma protein binding that takes place with these compounds.

Cocaine concentrations in saliva have been shown to correlate well with plasma concentrations and behavioral effects. Metabolites of cocaine are present in very low concentration in saliva. Inconsistencies in S:P ratios have been attributed to contamination of saliva following intranasal and smoked routes of administration. Peak cocaine concentrations in saliva increased from 428–1927 ng ml^{-1} after a single oral dose to between 15 000 and >500 000 ng ml^{-1} after smoking. Contamination of the oral cavity following these types of exposure precludes pharmacologic or behavioral interpretation of results. However, these results may indicate recent exposure to the drug, which could be of equal forensic importance.

Elevated THC concentration (50–1000 ng ml^{-1}) after smoking marijuana is also the result of oral contamination. Shortly after exposure, the concentration rapidly declines in a dose-dependent manner similar to that of plasma. However, some cannabinoids can be detected in saliva for longer periods than plasma, suggesting that some drugs are actually sequestered in the buccal cavity during smoking.

Semen

Analysis of semen for drugs of abuse has not been widely adopted. Privacy issues surrounding collection of this biofluid restrict its widespread use. Despite limited reports of drug detection in semen, it is known that certain drugs, such as cocaine, can influence the characteristics of the spermatozoa, causing decreased motility and anatomical abnormalities. It has also been shown that cocaine binds to sperm with high affinity, suggesting a unique mode of transport to the ovum. Although the concentration of drug in semen is not sufficient to elicit a response in a sexual partner, it has been suggested that insemination of drug-laden sperm into the egg could result in abnormal development.

Anatomy/physiology Semen is the viscous fluid released from the male upon ejaculation. The fluid contains spermatozoa as well as auxiliary sex gland secretions. Two seminal vesicles, the prostate gland and the bulbourethral glands contribute 95% of the gelatinous secretion. Seminiferous tubules of the testes contribute less than 5% of the seminal fluid volume, which is typically between 3 and 5 ml. Paired seminal vesicles, which are thin-walled, pear-shaped structures, secrete a thick, slightly alkaline fluid that mixes with the sperm as they pass into the ejaculatory ducts and urethra. These secretions, which are expelled when the seminal vesicles contract during orgasm, constitute about 60% of the seminal fluid volume. This secretion is rich in fructose, a sugar that stimulates the sperm to become mobile. The prostate gland, which sits just below the bladder, secretes a thin, milk-colored fluid, which accounts for about 30% of the total volume. This fluid helps activate sperm and maintain their motility. Bulbourethal glands located below the prostate produce mucus-like secretions that lubricate the terminal portion of the urethra, contributing about 5% of the total volume.

Sampling/analysis Following ejaculation, seminal fluid can be analyzed for drugs of abuse using techniques that are widely used for serum and other biofluids. Diluted seminal fluid may be used in immunoassay screening tests, and chromatographic analyses may be performed after extraction of the drug using conventional techniques. Semen pH typically ranges between 7.3 and 7.8, depending on the differential contribution of fluids. The overall alkalinity of the fluid helps protect spermatozoa from the acidic environment of the female reproductive tract.

Interesting/relevant findings Lipid solubility and pK_a of the drug play an important role in the transport of drugs of abuse into seminal fluid. Ion trapping may be responsible for the transport of certain drugs from the seminal plasma to the genitourinary tract. Drug ionization depends on the pH difference between plasma (pH 7.4) and prostatic fluid (pH 6.6); prostatic fluid can trap basic drugs in the prostate. In contrast, the vesicular fluid, which is more alkaline in nature, is likely to contain much lower concentrations of these drugs. Cocaine was detected in semen at concentrations typically 60–80% of those measured in plasma, independent of the route of administration. Following a 25 mg dose of cocaine, parent drug and benzoylecgonine concentrations in semen were 45 and 81 $\mu g\, g^{-1}$, respectively, at 1 h. One study involving opioid-maintained individuals indicated that methadone was present in semen at concentrations in excess of those found in blood. Amphetamine concentrations in semen were reported to correlate well with whole blood concentration following drug exposure. However, the disposition of drugs of abuse in this biofluid is widely variable due to the heterogeneous nature of the seminal fluid.

Sweat

A study performed by the Michigan State Department of Corrections evaluated sweat and urine as possible drug-detection matrices. Participants in the study were prisoners, subject to residential or electronic monitoring, who might have had access to street drugs. Sweat patches, which were applied and worn

for 7–14 days, were more effective indicators of drug use than repeated urinalysis every 3 days. Sweat analysis is becoming increasingly popular in drug compliance programs, such as rehabilitation, probation or parole, because it is a noninvasive, convenient means of specimen collection and requires less frequent testing.

Anatomy/physiology About 3 million tiny structures deep within the skin are responsible for the elimination of between 300 and 700 ml of fluid in a 24 h period. These glands secrete sweat, which is then transported through narrow passageways to the surface of the skin, whereupon evaporation has a cooling effect that helps maintain body temperature. The glands themselves are controlled by the autonomic nervous system, which is responsible for increased rates of secretion during times of anxiety or fear. Elimination of this body fluid, which occurs during normal breathing, is known as insensible sweat. Sensible sweat, which refers to perspiration that is actively excreted during stress, exercise or extreme temperature, may be eliminated at rates of 2–4 l h^{-1}. About half of the total volume of sweat is eliminated from the trunk of the body. The remaining fluid is lost from the legs or upper extremities and head in approximately equal amounts. The fluid consists of water (99%), sodium chloride, phosphate, protein, urea, ammonia and other waste products. The average pH of sweat is about 5.8, but increased flow rates increase the pH to between 6.1–6.7 in a manner analogous to saliva.

Sweat is produced by two types of gland, eccrine and apocrine. The former are coiled, tube-like structures that are placed throughout the skin. These glands, which are concentrated in the palms of the hands and soles of the feet, open up directly to the surface through tiny pores in the skin. Apocrine glands, which become active only after puberty, are large, deep glands in the axillae, pubic and mammary regions. These produce cellular material as well as sweat, which results in the secretion of a more viscous fluid with a characteristic odor. Apocrine glands often open up into hair follicles before reaching the surface of the skin. Contamination of the exterior surface of the hair has been demonstrated as a result of this biofluid exchange.

Parts of the body's surface are also bathed in sebum, an oily secretion of the sebaceous glands. Composed of keratin, fat and cellular debris, sebum mixes with sweat to form a moist, acidic film on the surface of the skin, which protects it from drying. This waxy lubricant, which consists mostly of fatty acids, maintains a pH of about 5 and is mildly harmful to bacteria and fungus, providing additional protection against infection.

Sampling Sweat is usually collected using an adhesive absorbent patch, which is placed on the surface of clean skin or by wiping the skin with a swab or gauze. Careful preparation of the skin is necessary before placement of a sweat patch to minimize external drug contamination or bacterial degradation of the drug once it has been retained. Occlusive sweat collection devices typically consist of an absorbent pad with an adhesive polyurethane exterior, similar to a waterproof bandage. These allow sweat to diffuse into the patch but prevent water or compounds from the environment from penetrating the device. In some instances, discomfort and inconvenience led to high rates of noncompliance among sweat patch users, resulting in tampering, loss of the patch or refusal to wear. Due to the relatively small volume (microliters) of insensible sweat secreted from a small absorbent area (typically 3 × 5 cm), patches are typically worn for several days on the outer portion of the upper arm or back. Sweating may be induced by occlusive wrapping or by diaphoretic treatment, but these efforts, which are more invasive, are not routinely used.

Occlusive bandages have been replaced with more advanced technology which improves user comfort. The PharmCheck™ sweat patch is a specimen collection device for nonvolatile and liquid components of sweat, including drugs of abuse. Nonvolatile components in the environment cannot penetrate the pad externally, and a semipermeable membrane that covers the absorbent pad allows oxygen, water and carbon dioxide to diffuse through. This helps maintain healthy skin, improves user comfort and increases the likelihood of compliance. Salts, solids and drugs that pass through the skin are trapped in the absorbent pad, where they are temporarily stored *in situ* until the patch is removed.

Most skin wipes and patches contain a mixture of sweat and sebum, both of which are secreted from the surface of the skin. Most reports of drugs of abuse in sweat actually refer to the collection of the mixed matrix. As yet, there have been relatively few reports of drugs of abuse in sebum alone. This biofluid is typically collected from the forehead, which is rich in sebaceous glands, as are the face and scalp. Unlike sweat, which is predominantly water, fat-soluble drugs can be sequestered in sebum due to the high lipid content of this fluid.

Analysis Surface secretions of sweat and sebum must be extracted before an initial drug screen. Drugs, which are present in the absorbent collection material, are generally extracted with alcohol, buffer or a combination of both. Recovery of commonly abused drugs during this process is usually high (80% or more) and limits of detection by GC-MS are

generally less than 1 ng per patch. The increased work necessary to analyze sweat patches is counteracted by the relative ease of sample collection and the ability to determine cumulative drug exposure over an extended period.

Interesting/relevant findings Quantitative analysis of drugs of abuse in sweat is rarely attempted because the volume of biofluid collected is generally uncertain. If an occlusive collection device is used, the volume of sweat can be estimated from the increased weight of the patch. Newer nonocclusive devices, which can be worn for longer periods, give no indication of specimen volume, but it has been suggested that the sample volume could be estimated indirectly from the concentration of sodium or lactate in the patch, both of which are excreted in sweat at relatively constant rates.

One of the principal advantages of sweat analysis is that it can provide information on long-term continuous exposure to drugs of abuse. Over a period of days, sweat saturates the absorbent pad, which retains the drugs. As most drugs are eliminated in the urine within 2–3 days of a single exposure, effective drug monitoring necessitates urinalysis every few days, which is inconvenient and labor intensive. Sweat patches, on the other hand, can be worn for as long as 2 weeks, are less likely to be adulterated and require drug testing to be performed only once, when the patch is removed.

Incorporation of drugs into sweat by transdermal migration or passive diffusion is favored by small, lipid-soluble compounds. Drugs that are highly bound to plasma proteins, such as benzodiazepines, are present in sweat at low concentrations. Basic drugs with high partition coefficients and pK_a values close to sweat appear to be maximally excreted. The patch itself operates as an ion trap towards weakly basic drugs that ionize as a result of the pH differential between plasma (pH 7.4) and sweat (pH 5–6). The predominant analyte in sweat appears to be the parent drug, which is generally less polar than the metabolites commonly found in urine. However, relative amounts of parent drug and metabolite vary between drugs and subjects, and this may be attributed to differences in skin enzymes or excretory pathways. Unlike many other biofluids, heroin, which is metabolized very rapidly by the body, is excreted in the sweat. Both heroin and its metabolite, 6-acetylmorphine, have been detected, both of which are conclusive markers of illicit opioid use.

Using a sweat patch, one-time exposure to cocaine can be detected for up to 7 days. Cocaine concentrations greater than 100 ng ml^{-1} were detected in sweat 72 h after a 2 mg kg^{-1} intranasal dose. Appearance of the drug in sweat occurs within 1–2 h of exposure and peaks within 24 h. Ion trapping may increase the detection window of certain basic drugs in sweat compared with in urine. In one study, methamphetamine was detected in sweat for up to 140 h after a 10 mg dose, compared with 96 h in urine. Individual skin wipes obtained from methamphetamine users indicated nanogram quantities of both methamphetamine (20–164 ng) and amphetamine (3–13 ng) on the surface of the skin.

Adulteration of patches is rare. Adulterants, which are commonly used for urinalysis, are difficult to apply without causing noticeable disturbance to the patch. Introduction beneath the adhesive layer using a hypodermic needle is possible, at the risk of causing substantial irritation to the skin as a result of the reactive or caustic nature of many adulterants. Once patches are sealed, the skin is free from further environmental drug contamination. However, studies have shown that external deposition of drug on the surface of the skin may result in detectable amounts of drug for several hours after normal hygiene practices have been performed. Cleaning the skin with alcohol before placement of a sweat patch may not sufficiently remove residual drug in these instances. *In situ* storage of excreted drugs in the sweat patch before removal makes drug stability in this biofluid an important consideration. Presence of drug metabolites in sweat may not reliably indicate *in vivo* drug exposure. Sweat contains esterases and other enzymes that may degrade or transform a drug on the surface of the skin. Hydrolysis of drugs such as cocaine and heroin to benzoylecgonine and 6-acetylmorphine, respectively, may take place inside the sweat patch.

Variation between subjects and environmental effects suggest that a correlation between sweat and blood concentration is unlikely, except perhaps under conditions of controlled sweating. Pharmacologic interpretation of drugs in sweat is not possible. The presence and quantity of drug present in a sweat sample can only be used to indicate exposure to that substance and is not indicative of impairment. However, this biofluid is unique, in as much as it offers convenient and noninvasive drug monitoring in drug compliance programs, which are becoming increasingly popular and necessary owing to increasing trends in drug use.

See also: **Drugs of Abuse**: Analysis.

Further Reading

Anderson KN (1998) *Mosby's Medical Dictionary*. St Louis: Mosby-Year Book.

Cone EJ and Jenkins AJ (1997) Saliva drug analysis.

Handbook of Therapeutic Drug Monitoring and Toxicology, pp 303–333. Boca Raton, FL: CRC Press.

Huestis MA and Cone EJ (1998) Alternative testing matrices.. *Drug Abuse Handbook*, pp 799–857. Boca Raton, FL: CRC Press.

Kidwell DA, Holland JC and Athanaselis S (1998) Testing for drugs of abuse in saliva and sweat. *Journal of Chromatography B* **713**:111–135.

Pichini S, Zuccaro P and Pacifici R (1994) Drugs in semen. *Clinical Pharmacokinetics* **26**:356–373.

Pons G, Rey E and Matheson I (1994) Excretion of psychoactive drugs into breast milk. *Clinical Pharmacokinetics* **27**:270–289.

Schramm W, Smith RH and Craig PA (1992) Drugs of abuse in saliva: a review. *Journal of Analytical Toxicology* **16**:1–9.

Spiehler V (1997) Detecting drugs of abuse in sweat. *Therapeutic Drug Monitoring and Toxicology* **18**:37–45.

Sunshine I and Sutliff JP (1997) Sweat it out. *Handbook of Therapeutic Drug Monitoring and Toxicology*, pp 253–264. Boca Raton, FL: CRC Press.

Classification, including Commercial Drugs

L A King, Forensic Science Service, London, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0484

Introduction

The term ‘drug abuse’ invites questions about what is a drug and what is the meaning of ‘abuse’. In a general sense, a drug is any substance without nutritional value which is capable of exerting a physiological or behavioral response in the subject. For the forensic drug examiner, as opposed to the toxicologist, this is much too broad. Following Paracelsus, we recognize all drugs as poisons, but the deliberate or accidental administration, including overdoses, of pharmaceutical medicines or the true poisons (e.g. carbon monoxide, cyanide, paraquat and arsenic) is not normally thought of as drug abuse. On the other hand, we should not conclude that the term is restricted to substances with medically useful properties. Most opinion asserts that cannabis, heroin and lysergide have little clinical value, yet they are widely abused. Alcohol and nicotine can lead to serious health and social problems, but there is much resistance in the wider community to even think of these substances as drugs.

For all practical forensic purposes, and for this discussion, drug abuse is largely concerned with those substances whose possession or supply is prohibited in law. We refer to them as the ‘scheduled’ or ‘controlled’ drugs. A few nonscheduled drugs are also of interest, but even here it is convenient to think of them as drugs which are potentially liable to be scheduled in the future. Expressions such as ‘illegal drug’ and ‘illicit drug’ are not always helpful. The former could include drugs which are perfectly legal when used in an appropriate context (e.g. morphine, benzodiazepines and other scheduled medicines), whereas ‘illicit’ refers to substances manufactured without licence and is not synonymous with ‘scheduled’. The use of labels such as ‘hard’ and ‘soft’ should also be avoided as they give the impression that some scheduled drugs are more acceptable than others. Although the expressions ‘abuse’ and ‘misuse’ are used interchangeably, the policy of the World Health Organization is to prefer the former if scheduled drugs are involved.

The legal classification of drugs is partly determined by their pharmacological properties. It is also convenient to group drugs according to their geographical origin, whether they are naturally occurring, semisynthetic or completely synthetic, their mode of administration, scale of abuse and physical form (e.g. powder, tablets, liquids).

In this article, drugs of abuse are described under three major headings:

- *Form and origin*
  - Natural plant material
  - Derived plant material
  - Semisynthetic drugs
  - Synthetic drugs
- *Pharmacological classification*
  - Narcotic analgesics
  - CNS stimulants
  - Hallucinogens
  - Hypnotics/tranquilizers
  - Miscellaneous
- *Legislative controls*
  - Scheduled drugs: international classification
  - The UN Single Convention on Narcotic Drugs, 1961
  - The UN Convention on Psychotropic Substances, 1971
  - National legislation
  - Nonscheduled drugs

Form and Origin

Drugs of abuse fall into four groups: natural plant material, derived plant material, semisynthetic drugs and true synthetic drugs.

Natural plant material

Herbal cannabis or marijuana typically consists of the dried flowering tops of female *Cannabis indica* and related species, but it may also include the dried leaves. This plant has a widespread distribution, but grows best in subtropical conditions. Plants grown outdoors in Northern Europe rarely produce a good yield of flowers. The active principle in cannabis is tetrahydrocannabinol (THC), much of which is found in the resin surrounding the bracts. Cannabis is normally mixed with tobacco and smoked, but can be ingested. The average 'reefer' cigarette contains around 200 mg of herbal cannabis. The mean THC content of good quality cannabis is around 5%. Improved seed varieties, which have been selected to produce higher THC yields, are associated with intensive indoor cultivation. Procedures such as artificial heating and lighting, control of 'day' length, hydroponic cultivation in nutrient solutions and propagation of cuttings of female plants not only lead to a high production of flowering material, but the THC content may be in excess of 20% in ideal cases. At the other end of the scale, cannabis is also cultivated under license in a number of European countries for the production of fiber. The THC content of this cannabis is less than 0.3%.

The leaves of the coca plant (*Erythroxylum coca*), which grows on the Andean ridge in South America, have long been chewed by local people as a stimulant Coca leaves are also made into a commercially available coca tea, but this product is rarely seen outside South America. The active principle is cocaine.

Mescal buttons from the peyote cactus (*Lophophora williamsii*) originate from Central America and have had a long use in religious ceremonies. The active constituent is mescaline, an hallucinogenic phenethylamine. The peyote cactus is a slow-growing plant, which is rarely cultivated on a commercial scale. Mescaline has largely been superseded by synthetic phenethylamines (see below).

'Magic' mushrooms (*Psilocybe semilanceata* and related species) are another ubiquitous group. They contain the hallucinogenic tryptamines, psilocin, and its phosphate ester, psilocybin. The related hallucinogen, dimethyltryptamine, is a constituent of various tropical plants.

Khat or qat (*Catha edulis*) is a plant indigenous to Northeast Africa. The fresh leaves are chewed to extract both cathinone and cathine. These substances are stimulants closely related to ephedrine. Abuse of khat in Western countries is mostly confined to Ethiopian and Somalian communities.

Derived plant material

Cannabis resin (produced by collecting and compressing the resinous matter surrounding female cannabis flowers) has a typical THC content of 5%. Hash oil (a solvent extract of herbal cannabis or cannabis resin) may contain 30–40% THC.

Cocaine comprises around 1% of coca leaves. It is chemically extracted and manufactured into cocaine hydrochloride. This is a white powder which is commonly snorted. Some cocaine is reconverted to the free base. Known as 'crack', it is extremely addictive. Unlike the hydrochloride salt, it can be smoked.

Opium is also smoked, but abuse is uncommon in Western countries. It is the dried latex which exudes from the capsules of the opium poppy (*Papaver somniferum*), and contains around 10% morphine as the main active ingredient. Poppy straw concentrates, made by solvent extracting the dried capsules, are sometimes injected – a practice seen in Eastern Europe.

Semisynthetic drugs

Heroin is the most common example of a semisynthetic drug. It is produced by the acetylation of crude morphine obtained from opium. Until the late 1970s, nearly all of the heroin consumed in Europe came from Southeast Asia. Sometimes known as 'Chinese heroin', it was a white powder consisting of diamorphine hydrochloride and minor amounts of other opium alkaloids, but adulterants were unusual. However, over the past 20 years, most of the heroin seized in Europe has originated in Southwest Asia, an area centred on Pakistan, Afghanistan and Turkey. This material is a much cruder product, typically a brown powder containing around 45% diamorphine and variable amounts of other opium alkaloids (e.g. monoacetylmorphine, noscapine, papaverine and acetylcodeine). Adulterants are normally present, the most frequently reported being caffeine and paracetamol. It is generally accepted that most of these cutting agents are added to heroin at the time of manufacture. The illicit heroin used in North America traditionally originated from Southeast Asia, but in recent years Columbia has emerged as a major heroin supplier.

Synthetic drugs

Amphetamine was first synthesized over 100 years ago, but its stimulant properties were not recognized until much later. It is now rarely prescribed as a medicine, where one of its few uses is in the treatment of narcolepsy. In Europe, amphetamine is the second

most commonly abused drug after cannabis. Most amphetamine syntheses continue to start with phenyl-2-propanone (benzyl methyl ketone). Despite international trade controls deriving from the United Nations (UN) 1988 Convention, illicit supplies of this and other precursors appear to be readily available on the black market. In North America and the Far East, methamphetamine is more common. This has traditionally been made from ephedrine, but trade controls have caused a shift to pseudoephedrine. More recently, phenylpropanolamine has been used in the USA to produce amphetamine. Both amphetamine and methamphetamine are found as white or off-white powders. Typical cutting agents are caffeine and sugars such as glucose. In Europe, amphetamine is occasionally seen in tableted form.

The ring-substituted amphetamines are commonly known as the 'Ecstasy' drugs. The prototypical member of this family is 3,4-methylenedioxymethamphetamine (MDMA). First synthesized in the early part of the twentieth century, abuse did not become widespread until the 1980s. Both amphetamine and MDMA are derived from the phenethylamine molecule. In the past 10 years, dozens of designer drugs based on variously substituted phenethylamine have appeared in Europe and the USA. These drugs are invariably produced in the form of white well-made tablets, often bearing a characteristic design (logo) and usually around 10 mm in diameter. The MDMA content of tablets is typically 80–90 mg. Lactose is a common excipient (filler) in tablets.

Lysergide (LSD) is generally thought of as a purely synthetic material, but routes of manufacture usually start from ergotamine, a natural substance produced by the microorganism *Claviceps purpurea*. Until the mid-1970s, LSD was produced in small (approximately 2×2 mm) tablets known as microdots. For the past 20 years, paper squares of around 7×7 mm have been the common dosage form. These squares are usually printed with a coloured design featuring cartoon characters, symbols or drug-related motifs; the lysergide content averages 50 μg.

Apart from lysergide and dimethyltryptamine (DMT), few synthetic drugs based on the tryptamine molecule have become popular, even though the synthesis and properties of many have been described in recent popular literature. A factor that is likely to limit the wider abuse of hallucinogenic tryptamines is their inactivity when taken orally. Most need to be smoked, injected or mixed with an 'activator' to inhibit metabolic destruction.

Methaqualone, a drug formerly used as a hypnotic, but now produced illicitly, has remained a popular drug of abuse in South Africa. It is often smoked mixed with cannabis.

Pharmacological Classification

The legal classification of an abused drug is determined by its propensity to harm the health of the individual or produce a risk to society. In large measure, these factors are governed by the pharmacological properties of the substance, particularly its toxicity and ability to produce dependence (a term used in preference to addiction). Most abused drugs fall into a few well-defined pharmacological categories, namely narcotic analgesics, central nervous system (CNS) stimulants, hallucinogens and hypnotics/tranquilizers. A small number fall into a miscellaneous or mixed function group.

Narcotic analgesics

These are drugs which interact with those receptors in the brain responsible for the transmission of and response to pain. They may be differentiated from the peripheral analgesics (e.g. aspirin), which have no abuse potential. The classical narcotic analgesics are the opium alkaloids (principally morphine) and its semisynthetic derivatives (e.g. diamorphine – the main active principle in heroin, codeine, buprenorphine). In the first half of the twentieth century, a large group of entirely synthetic narcotic analgesics was developed, including methadone, pethidine (meperidine) and fentanyl. Abuse of narcotic analgesics is responsible, by far, for most drug-related mortality and morbidity.

CNS stimulants

Although the narcotic analgesics may cause the greatest damage to society and the individual, the CNS stimulants are probably the most widely abused. These drugs have a close structural relationship to neurotransmitters such as dopamine and noradrenaline. They are believed to act on receptors in the brain and other tissues, causing an increase in the levels of these neurotransmitters. This results in a rise in blood pressure, increased mental alertness and wakefulness, reduction of physical fatigue and appetite. Prolonged use can lead to psychosis. In developed countries, amphetamine, methamphetamine and cocaine are the commonly encountered examples. In some countries, there is abuse of medicinal drugs nominally intended for the treatment of narcolepsy, weight reduction or hyperactivity attention disorder (e.g. diethylpropion, pemoline, methylphenidate).

Hallucinogens

Naturally-occurring hallucinogens have been used in Central and South America since historical times. The most common of these are mescaline, dimethyltryptamine and other ring-substituted tryptamines

such as psilocin. These drugs produce visual and other sensory hallucinations. Synthetic hallucinogens are typified by lysergide. It is one of the most potent drugs to act on the CNS. Although first produced over 50 years ago, serious abuse was uncommon before the 1960s. This era also saw the appearance of other clandestine hallucinogens such as DOB (brolamfetamine, 4-bromo-2,5-dimethoxy-α-methylphenethylamine). More recently, other hallucinogenic phenethylamines have received wide publicity. These drugs, sometimes known as psychedelics, are thought to act by interfering with serotonin receptors in the brain. As with the stimulant drugs, there is a close relationship between the chemical structure of hallucinogenic drugs and serotonin (5-hydroxy-*N*,*N*-dimethyltryptamine).

Hypnotics/tranquilizers

The barbiturates were once the most commonly used hypnotic drugs, but are now rarely prescribed or abused. Their CNS depressant properties led to many deliberate and accidental overdoses. They have been replaced by the benzodiazepines – a group which spans a range of both hypnotic (sleep-inducing) and tranquilizing properties. Benzodiazepine tranquilizers are typified by diazepam and lorazepam. Although these drugs may often be found in impaired vehicle drivers, abuse is largely restricted to the hypnotic members (e.g. flunitrazepam and temazepam). However, the distinction between the two types is largely based on their potency and duration of action.

Miscellaneous

The so-called 'Ecstasy' drugs are sometimes described as hallucinogenic, but their effects are unlike those of LSD. Although MDMA and its congeners may show some stimulant properties, they have been described as falling into the novel pharmacological categories of 'entactogens' and 'empathogens'. The pharmacological effects of cannabis are fairly diverse, as may be expected from a complex plant product. Effects include euphoria, sedation, analgesia and hallucinations. Most abused solvents fall into the category of simple anesthetics, although substance-specific effects have been reported. Anabolic steroids are abused for their ability to increase lean body weight, strength and overall physical fitness.

Legislative Controls: Scheduled Drugs: International Classification

UN Single Convention on Narcotic Drugs, 1961

Plant-based drugs, such as opium, cannabis and cocaine, have been used for thousands of years, but the concept of drug abuse is much more recent. As a consequence of social change, the problem has inevitably arisen from a situation where the drug in question was initially accepted, at least in certain populations, if not actively promoted by commercial interests. This was the position with opium towards the end of the nineteenth century. Following the meeting of the Shanghai Opium Commission in 1909, trade was curtailed and eventually replaced by legislative sanctions against supply and use. Cocaine was once a permitted additive in Coca-Cola, whereas heroin (diacetylmorphine) was marketed as a treatment for opium addiction.

The major plant-based drugs have now been variously controlled for many years, but the modern era of international legislation starts with the UN Single Convention on Narcotic Drugs in 1961. Member States which are signatories to the Convention will have established the principles in domestic legislation. The Convention maintains a strong emphasis on these same plant-based drugs, setting out rules for their cultivation, manufacture and trade. However, the scope of control was widened to include over 100 mostly synthetic substances, the great majority of which can be described as narcotic analgesics; few of these are now used clinically or ever abused. The drugs are set out in four Schedules. Most are found in Schedule I, the category with the greatest restrictions. The more commonly encountered drugs in forensic casework, which are included in the 1961 Convention, are set out in **Table 1**.

UN Convention of Psychotropic Substances, 1971

During the 1960s, the abuse of stimulants, hallucinogens and related drugs became a problem in many Western societies. These drugs were often abused in the form of pharmaceutical preparations. Just as with the plant-based drugs in a previous era, the supply of stimulants was promoted by commercial concerns as aids to weight control and relief from fatigue. Even hallucinogens, such as LSD, were legitimized by their use in psychiatry. The 1971 UN Convention on Psychotropic Substances was intended to deal with this phenomenon. Again, the drugs for control were set out in four Schedules. Whereas a forensic drugs examiner is unlikely to encounter the great majority of substances in the 1961 Convention, the psychotropic drugs in the 1971 Convention will be more familiar. There are over 100 substances listed; **Table 2** shows the more commonly encountered drugs in the four Schedules. Unlike the 1961 Convention, there is no overarching control of isomers. This leads to a situation in which a generic term such as amphetamine (meaning both the '−' and the '+' enantiomers)

Table 1 The more commonly encountered drugs of abuse listed in the 1961 UN Single Convention on Narcotic Drugs

Schedule I
Cannabis and cannabis resin and extracts and tinctures of cannabis
Coca leaf
Cocaine
Concentrate of poppy straw
Dextromoramide
Dipipanone
Heroin (diacetylmorphine)
Methadone
Morphine
Opium
Pethidine

The isomers, esters and ethers (unless appearing in another Schedule) and the salts of the drugs listed in this Schedule

Schedule II
Codeine
Dihydrocodeine

The isomers and the salts of the drugs listed in this Schedule

Schedule III
This is concerned with what may be termed 'low-dose' preparations of drugs already listed in Schedules I or II. For example, it includes: codeine or dihydrocodeine when compounded with one or more other ingredients and containing not more than 100 mg of the drug per dosage unit and with a concentration of not more than 2.5% in undivided preparations; cocaine preparations containing less than 0.1% cocaine; and morphine preparations containing less than 0.2% morphine

Schedule IV
This is a list of those Schedule I substances which, unlike most of the entries in Schedule I, are considered to be of limited medicinal value and should be restricted to research use. It includes, among others, cannabis and cannabis resin and heroin

Table 2 The more commonly encountered drugs of abuse listed in the 1971 UN Convention on Psychotropic Substances

Schedule I
Brolamphetamine (DOB)
Cathinone
Dimethyltryptamine (DMT)
N-Ethyltenamphetamine (MDE)
N-Hydroxytenamphetamine (*N*-OH MDA)
Lysergide (LSD)
MDMA
Mescaline
Psilocin
Psilocybin
Tenamphetamine (MDA)

Schedule II
Amphetamine
Dexamphetamine
Dronabinol
(+)-Methamphetamine
Methamphetamine racemate
Methaqualone
Methylphenidate
Phencyclidine
Secobarbitone

Schedule III
Amobarbital
Buprenorphine
Cathine
Pentobarbital

Schedule IV
Diazepam and many other benzodiazepines
Diethylpropion
Pemoline
Phenobarbitone

The salts of the substances listed in all four Schedules

sits alongside dexamphetamine (i.e. the '+' enantiomer of amphetamine).

National legislation

Nearly all States limit their domestic controls to those substances listed in the 1961 and 1971 UN Conventions. A few countries have chosen to control a wider range of drugs. For example, certain anabolic steroids are controlled in the UK and the USA, although the substances named are not identical in the two countries. In the UK, 48 anabolic steroids are listed specifically as Class C drugs (the lowest category), and generic legislation covers certain derivatives of 17-hydroxyandrostan-3-one or 17-hydroxyestran-3-one as well as the esters or ethers of the named steroids. In the UK, the more commonly encountered controlled steroids are methandienone, nandrolone, oxymetholone, stanozolol, and testosterone and its esters. Further anabolic nonsteroidal compounds are also controlled, i.e. human chorionic gonadotrophin (hCG), clenbuterol, nonhuman chorionic gonadotrophin, somatotropin, somatrem and somatropin.

Starting in the late 1970s, the UK introduced a number of generic definitions into the Misuse of Drugs Act of 1971. Thus, apart from derivatives of certain steroids as noted above, suitably substituted derivatives of tryptamine, phenethylamine, fentanyl, pethidine and barbituric acid also became controlled. A broadly similar approach is used in New Zealand. In the USA, the problem of so-called 'designer drugs' has been effectively dealt with by the Controlled Substances Analogue Enforcement Act of 1986.

In the UK legislation, apart from a few exceptions, no distinction is made between the isomers of listed drugs. Thus the Misuse of Drugs Act lists amphetamine, but not dexamphetamine, which is considered redundant. By contrast, in the USA, offenses involving different isomers of, for example, methamphetamine may attract different penalties.

Some countries have scheduled specific substances either for legislative convenience or because the drugs are deemed to be a local problem. An example of the former is the inclusion of ephedrine as a controlled drug in the Republic of Ireland. (This arose from a desire to consolidate into their Misuse of Drugs Act of 1977 the provisions of the UN 1988 Convention Against Illicit Traffic in Narcotic Drugs and Psychotropic Substances.) An example of the latter is the control of the khat plant in the USA and a number of European countries.

Nonscheduled Drugs

Apart from drugs under international or domestic control, there is a further small group of substances that are recognized as causing personal and social problems, particularly in Europe and North America. These are possible candidates for control, but as discussed earlier should be distinguished from 'socially acceptable' substances such as alcohol, nicotine and caffeine.

Ketamine and gamma-hydroxybutyrate (GHB) are both under investigation in the UK and USA as candidates for control. Ketamine is used in veterinary and some human surgery as an anesthetic. Only injection solutions are licensed, whereas abusers ingest powders and tablets, sometimes mixed with a stimulant such as ephedrine to mimic the effects of MDMA. GHB is licensed for use in some countries as a hypnotic, but it is also abused. Not only is GHB readily made from its precursor (gamma-butyrolactone), but that precursor is widely used as an industrial solvent which is metabolically converted to GHB.

Abuse of solvents by inhalation is arguably a more serious problem, which leads to many fatalities. However, hydrocarbons, such as butane and toluene, are so readily available that effective control would prove difficult to achieve. The alkyl nitrites, which cause peripheral vasodilation, represent a particular type of solvent abuse, but again their control would present practical problems.

Finally, there are numerous herbal drugs which are abused for their stimulant or hallucinogenic properties. Not only is this somewhat of a fringe activity, but most legislatures are reluctant to control a wider range of plant material.

See also: **Drugs of Abuse**: Designer Drugs. **Pharmacology**.

Further Reading

Advisory Council on the Misuse of Drugs (1998) *Drug Misuse and the Environment*. London: Stationery Office.

British Medical Association (1997) *The Misuse of Drugs*. Amsterdam: Harwood.

European Monitoring Centre for Drugs and Drug Addiction (1998) *Annual Report on the State of the Drugs Problem in the European Union*. Luxembourg: Office for Official Publications of the European Communities.

Fortson R (1996) *The Law on the Misuse of Drugs and Drug Trafficking Offences*, 3rd edn. London: Sweet and Maxwell.

Hardman J (eds) (1996) *Goodman and Gilman's Pharmacological Basis of Therapeutics*. New York: McGraw-Hill.

Henry J (1994) *The British Medical Association New Guide to Medicines and Drugs*. London: Dorling Kindersley.

Institute for the Study of Drug Dependence (1999) *Drug Abuse Briefing: A Guide to the Non-Medical Use of Drugs in Britain*, 7th edn. London: ISDD.

Julien RM (1995) *A Primer of Drug Action*, 7th edn. New York: Freeman.

Karch SB (eds) (1998) *Drug Abuse Handbook*. London: CRC Press.

Ramsay M and Spiller J (1997) *Drug Misuse Declared in 1996: Latest Results of the British Crime Survey*. London: Home Office.

Redda KK, Walker CA and Barnett G (ed.) (1989) *Cocaine, Marijuana, Designer Drugs, Chemistry, Pharmacology and Behavior*. Boca Raton, FL: CRC Press.

Stockley D (1992) *Drug Warning*. London: Optima.

United Nations (1993) *Multilingual Dictionary of Narcotic Drugs and Psychotropic Substances under International Control*. New York: United Nations.

United Nations International Drug Control Programme (1997) *World Drug Report*. Oxford: Oxford University Press.

Webber D (ed.) (1998) *Controlled Drugs: A Handbook for the Legal Profession*. London: Institute for the Study of Drug Dependence.

Designer Drugs

J Hartelius, Stockholm, Sweden

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0495

Emergence of the Concept

The concept of 'designer drugs' gained attention in the 1980s as a new expression for describing a number of recent arrivals among the psychoactive

substances on the illegal drug markets. When introduced for intoxicating purposes, the substances in many cases escaped formal drug control, such as the Controlled Substances Act (US), the Misuse of Drugs Act (UK) or the Narcotics Control Act (Sweden); nor were they regulated under the United Nations International Drug Conventions (Single Convention on Narcotic Drugs, 1961; Convention on Psychotropic Substances, 1971). Thus, they were rapidly perceived by public authorities to constitute a challenge to traditional drug control.

The concept of designer drugs is generally attributed to Dr Gary L. Henderson of the University of California at Davis, who introduced the concept to describe new, untested, legal synthetic drugs mimicking the effects of illicit narcotics, hallucinogens, stimulants and depressants. The concept has, however, no established scientific or codified judicial definition. Instead, several definitions have been advanced: most converge on the following criteria, stating that designer drugs are:

1. psychoactive substances;
2. synthesized from readily available precursor chemicals;
3. marketed under attractive 'trade marks' (fantasy names);
4. not subjected to legal control as narcotic drugs or psychotropic substances, at least not at the time of their introduction on the nonmedical drug market.

Sometimes the following criteria are added:

5. being produced mainly or exclusively in clandestine laboratories;
6. having little or no established medical use.

In common parlance, a designer drug can be said to be a synthetic psychoactive substance having pharmacologic effects similar to controlled substances and initially remaining outside traditional drug control.

Sometimes, these drugs are referred to as analogs, analog drugs or homologs, to stress the analogy or homology of their pharmacologic properties (and chemical structure) to well-known controlled substances.

In media reporting, the concept also has a connotation of novelty: the drugs are seen as modern creations in the same sense as designer clothes in fashion.

Some designer drugs have nevertheless been known for decades: the process for synthesizing 3,4-methylenedioxymethamphetamine (MDMA) was patented in 1914.

The demarcation of the substances regarded as designer drugs is not universally accepted. Sometimes 'old timers', such as methamphetamine (widely abused in Japan in the 1950s), are included. The concept is fluid and dynamic, mirroring a continuous search for new psychoactive substances. The pioneer in this field is the American pharmacologist and chemist Alexander Shulgin (b. 1925), who for decades has been synthesizing and investigating the effects of hundreds of such substances.

Some designer drugs have also been studied experimentally in cult-like settings in religiously flavored attempts to experience altered states of human consciousness.

Issues of Forensic Interest

Designer drugs raise at least four issues of forensic interest:

1. Determination of the chemical composition of new substances on the illegal drug market, such as when 'Hog' in 1968 was shown to be phencyclidine (PCP).
2. Problems of detecting very low concentrations (p.p.b. range) of a substance in body fluids in cases of severe intoxication or fatal poisoning. Sometimes the health hazard may come from impurities, such as 1-methyl-4-phenyl-1,2,5,6-tetrahydropyridine (MPTP) in the synthesis of 1-methyl-4-phenyl-4-piperidyl propionate (MPPP) as a heroin substitute resulting in severe parkinsonism (the worst cases known as 'frozen addicts'). The drugs have also be used secretly to intoxicate or sedate a person for aims of sexual abuse or robbery. The detection problem is aggravated by the fact that new substances may not be fully investigated in respect of metabolites and routine methods of detection.
3. Determination of the particular substance used in a case of individual drug intoxication, especially when police diagnostic methods (such as the American drug recognition expert (DRE) wayside procedures) are used, as some designer drugs may not be controlled substances within the scope of domestic drug legislation.
4. Strategies to be used for controlling new drugs, e.g. drug scheduling, when a large number of new and unknown substances appear on the market. The number of designer drugs that can be produced in clandestine laboratories has been estimated to be from 1000 upwards, depending on the competence and resources of the chemist.

Major Types of Designer Drugs

Designer drugs can be classified, in the same way as other psychoactive drugs, into depressants, stimulants, hallucinogens, etc. More commonly, they are

classified according to their basic chemical structure, such as fentanyls or phenetylamines. The most important contemporary designer drugs are to be found among the following classes.

Fentanyls

These have the same major properties as opiate narcotics, and they are used in medicine mainly as anesthetics or analgesics (**Fig. 1**). They are, however, much more potent; some varieties are estimated to be 20–4000 times as potent as heroin. They may produce intoxication even in the microgram range. One gram (1 g) may be sufficient for 10 000 doses. Because of their extreme potency, fentanyls may be difficult to dilute for street sale as intoxicants. Their clandestine synthesis may be extremely profitable: precursors costing USD 200 may produce drugs at a value of USD 2 000 000 on the illicit market. Alphamethylfentanyl, diluted and sold under the name 'China White' to mimic traditional heroin of extreme purity, was among the first designer drugs to appear on the market (1979) and rapidly caused a number of overdose deaths. The number of drugs in this class is estimated to be around 1400, and the total number of designer opiates 4000.

Methcathinone analogs

The khat (kath, qat) plant (*Catha edulis*) contains the central nervous system stimulants cathin (norpseudoephedrine) and cathinone. They are both under international control as psychotropic substances, but not the plant *per se*. Methcathinone (ephedrone) is a synthetic analog of cathinone, and it can easily be produced in a 'kitchen laboratory' (**Fig. 2**). It is highly dependence- and tolerance-forming, and it is often abused in 'binges' with intensely repeated administration for several days. Its main effects include severe anorexia, paranoia and psychosis. Homologs of methcatinone, some of which are not under formal drug control, are considered by some to represent a potentially new series of designer drugs yet to be introduced to the nonmedical market. The number of drugs in this class is estimated to be around 10.

Phencyclidines

These have both anesthetic and hallucinogenic properties. PCP was originally developed in the 1950s as an anesthetic (Sernyl), but it was withdrawn after it was found to produce severe hallucinations in patients (**Fig. 3**). Ketamine (Ketalar, Ketanest) is still used as an anesthetic, but it is also abused for its alleged ability to produce a meditative state. Drugs of this type may cause depression, psychosis, personality changes and overdose deaths. The confounding of anesthetic and hallucinogenic properties may make the abuser very difficult to handle in confrontations with the police. The number of drugs in this class is estimated to be 50.

Phenetylamines (PEAs)

These comprise a large group, including such highly diverse substances as amphetamine and mescaline. These drugs can be divided into two subgroups: mainly euphoriant or mainly hallucinogenic. Among the euphoriant PEAs, the most widely known and abused one is MDMA (this was the original 'Ecstasy' drug, but this name has also been applied to new varieties such as 3,4-methylenedioxyamphetamine (MDA) and *N*-ethyl-3,4-methylenedioxyethylamphetamine (MDEA)) (**Fig. 4**). It has become especially popular among young people at 'rave' and 'techno' parties. Sometimes these drugs are labeled 'entactogens' for their alleged ability to increase sensitivity to

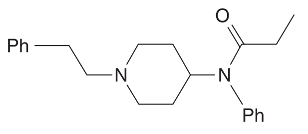


Figure 1 Fentanyl. Reprinted from Valter K, Arrizabalaga P (1998) *Designer Drugs Directory* with permission from Elsevier Science.

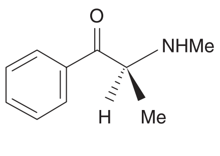


Figure 2 Methcathinone. Reprinted from Valter K, Arrizabalaga P (1998) *Designer Drugs Directory* with permission from Elsevier Science.

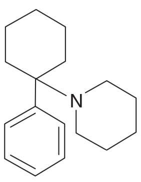


Figure 3 Phencyclidine. Reprinted from Valter K, Arrizabalaga P (1998) *Designer Drugs Directory* with permission from Elsevier Science.

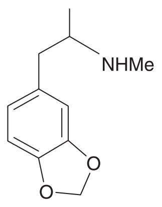


Figure 4 MDMA. Reprinted from Valter K, Arrizabalaga P (1998) *Designer Drugs Directory* with permission from Elsevier Science.

touch, or 'empathicogens', for their alleged ability to create empathy, especially before sexual encounters. They may cause severe neurotoxic reactions due to serotonin depletion (e.g. depression, suicidal behavior and personality changes) and, when taken in connection with physical exercise such as dancing, even rapid fatalities. The predominantly hallucinogenic (psychotomimetic) PEAs include e.g. DOB (4-Bromo-2,5-dimethoxyamphetamine), DOM (2,5-Dimethoxyamphetamine), and TMA (3,4,5-Trimethoxyamphetamine). These drugs may cause or trigger psychotic reactions. The total number of drugs in this class is estimated to be several hundred.

Piperidine analogs

Piperidine analogs are mainly MPPP, MPTP, OPPPP (1-(3-Oxo-3-phenylpropyl)-4-phenyl-4-piperidinol propionate) and PEPAP (1-Phenethyl-4-phenyl-4-piperidol acetate (ester)) (**Fig. 5**). They have both euphoriant and analgesic properties. They may cause, for example, convulsions and respiratory depression. MPPP is a powerful analgesic but it has never been used in clinical medicine. MPTP has been identified as the causative agent of the frozen addicts syndrome. The number of potential drugs in this class has not been estimated.

Tryptamines

Tryptamines (indolealkylamines) include drugs such as dimethyltryptamine, harmaline, lysergide (LSD) and psilocybine. They produce a variety of reactions, including hallucinations and tremor. LSD was the center of a psychedelic cult in the 1960s, being used for its alleged 'mind-expanding properties'. The number of LSD analogs is estimated to be 10 and the number of psychotomimetic indolealkylamines is estimated to be several hundred.

Trends and Control Issues

The modern designer drugs represent a new phase in drug synthesis and drug abuse. They have caused fundamental problems for contemporary administrative drug control, forensic investigation and medical services. In reviewing the problem, one author wrote that 'Designer drugs are subverting the black-market drug trade, undermining and subverting law enforcement activities, and changing our basic understanding of drugs, their risks and the marketplace itself.'

As only a small number of the thousands of varieties of designer drugs have appeared on the nonmedical market, it is likely that many more will be introduced, tested, abused and, ultimately, administratively controlled.

Over the next years, the emergence of new designer drugs will raise important control issues, such as the criteria and procedures to be used for extending drug control to new substances. This may call for a new definition of drugs (controlled substances), not solely based on administrative drug scheduling but also on the pharmacological properties of the new drugs.

Designer drugs will also raise complicated scientific and medical issues, such as new methods for identifying hazardous substances and treating severe intoxications. They may open new vistas for brain research.

A substantial increase in clandestine production of designer drugs may change the structure of the illegal drug market, making some established forms of production (e.g. heroin from opium poppy) obsolete.

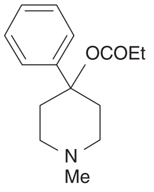


Figure 5 MPPP. Reprinted from Valter K, Arrizabalaga P (1998) *Designer Drugs Directory* with permission from Elsevier Science.

See also: **Pharmacology**.

Further Reading

Dal Cason TA (1997) The characterization of some 3,4-methylenedioxycathinone (MDCATH) homologs. *Forensic Science International* 87:9–53.

Holmstedt B (1969) Gas chromatography–mass spectrometry as a tool to elucidate the structure of unknown psycho-active drugs. In: Sjöquist F and Tottie M (eds) *Abuse of Central Stimulants*, pp. 357–373. Stockholm: Almquist and Wiksell.

Kalix P (1984) The pharmacology of khat. *General Pharmacology* 15:179–188.

Klein M, Sapienza F, McClain H and Khan I (1989) *Clandestinely Produced Drugs, Analogues and Precursors*. Proceedings of an International Conference, 8–11 September 1987, Rabat, Morocco. Washington, DC: WHO/Drug Enforcement Administration.

Kirsch MM (1986) *Designer Drugs*. Minneapolis: CompCare.

Shulgin A and Shulgin A (1991) *PiHKAL – A Chemical Love Story*. Berkeley, CA: Transform Press.

Shulgin A and Shulgin A (1997) *TiHKAL – The Continuation*. Berkeley, CA: Transform Press.

Valter K and Arrizabalaga P (1998) *Designer Drugs Directory*. Lausanne: Elsevier.

Drugs and Driving

S J Heishman, NIDA Addiction Research Center, Baltimore, MD, USA
S B Karch, City and County of San Francisco Hall of Justice, San Francisco, CA, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0490

Introduction

Motor vehicle accidents are the leading cause of death in the United States for people aged 1–34 years, and ethanol is a factor in nearly half the traffic fatalities each year. Studies investigating the prevalence rate of drugs other than ethanol in fatally injured drivers have reported varied results, ranging from 6 to 37%. Among individuals stopped for reckless driving who were judged to be clinically intoxicated, urine drug testing indicated 85% were positive for cannabinoids, cocaine metabolites, or both.

These relatively high prevalence rates reinforce the general assumption that psychoactive drugs are capable of impairing driving. Drug prevalence rates do not imply impaired driving; however, because of the reliability with which certain drugs degrade psychomotor and cognitive performance in the laboratory, it is highly likely that many drug-related vehicular accidents and driving under the influence/driving while impaired/intoxicated arrests involve impaired behaviors critical for safe driving. There are some studies that have investigated the effects of drugs on actual driving; however, the majority of our knowledge concerning drugs and driving is inferred from laboratory studies using driving simulators or tasks that are assumed to measure critical aspects of driving skills such as reaction time, divided attention, and vigilance. The information presented in the following section emphasizes data from studies of actual driving on closed courses or highways, but also includes pertinent data from laboratory studies examining the effects of drugs on human performance.

Specific Drugs and Driving

Psychomotor stimulants (cocaine, *d*-amphetamine, and methamphetamine)

The psychomotor stimulants, cocaine, *d*-amphetamine, and methamphetamine, share a similar psychopharmacological profile, such that low to moderate acute doses of both drugs increase positive mood, energy, and alertness in nontolerant individuals. It is well known that amphetamines can increase the ability to sustain attention over prolonged periods of time when performing monotonous tasks. Amphetamines or cocaine have also been shown to improve performance on auditory and visual reaction time tests; the digit symbol substitution test (DSST), a test of psychomotor skills and attention; and on tests of selective and divided attention.

Although numerous studies have shown that psychostimulants can enhance performance on tasks that appear related to driving (e.g. reaction time, motor coordination), no studies that directly assessed the effects of amphetamines in actual or simulated driving situations were found in recent literature searches. An indication of the incidence of stimulant use among truck drivers was recently reported. Of 317 drivers who volunteered urine or blood samples, 5% were positive for prescription amphetamines (amphetamine, methamphetamine, phentermine) and 12% were positive for over-the-counter sympathomimetic drugs, such as phenylpropanolamine, ephedrine and pseudoephedrine.

Chronic use of psychostimulants can result in a drug dependence syndrome, such that abrupt cessation of drug use can result in depression or fatigue. When stimulants are taken chronically in large doses, a toxic psychosis can develop that is characterized by vivid auditory and visual hallucinations, paranoid delusions, and disordered thinking, and is often indistinguishable from paranoid schizophrenia. Thus, although occasional use of low to moderate doses of amphetamines or cocaine can enhance performance, chronic use is not compatible with situations requiring clear thinking and coordination, such as driving a motor vehicle.

Cold and allergy medications (antihistamines)

All first-generation antihistamines produce sedation and have been shown to impair psychomotor and cognitive performance as well as on-road driving. For example, in a test of highway driving, diphenhydramine (50 mg) significantly impaired the ability to maintain steady lane position and to follow speed changes in a lead car. The second generation of antihistamines are less lipophilic than the previous generation and thus cross the blood–brain barrier less readily, which accounts for the lower levels of sedation observed with the newer drugs. Terfenadine was the first second-generation antihistamine to be approved by the Food and Drug Administration and marketed in the United States. In general, studies have found that terfenadine, in acute doses that exceeded the therapeutic dose (60 mg, b.i.d.), did not impair psychomotor and attentional performance as measured in the laboratory and in on-road driving tests. However, when terfenadine (120 mg, b.i.d.) was

administered for 3 days, driving impairment was observed. Similar results have been obtained with other second-generation antihistamines, such as loratadine, cetirizine, and mizolastine, when therapeutic doses are exceeded. Thus, the second-generation antihistamines produce less sedation than first-generation compounds and are generally safe at therapeutic doses. However, when therapeutic doses are exceeded, the so-called 'nonsedating' antihistamines are capable of producing sedation, which could result in impaired driving.

Sedative-hypnotics (benzodiazepines)

Because of their sedative effects, benzodiazepines generally impair psychomotor performance in non-tolerant individuals. Not surprisingly, benzodiazepines are associated in a dose-dependent manner with a significantly increased risk for traffic accidents. In studies of on-road driving, diazepam (5 mg, t.i.d.), lorazepam (0.5 mg, t.i.d.), and lorazepam (2 mg, b.i.d.) increased lateral lane position and slowed response time to a lead car's change in speed. Several benzodiazepines (oxazepam 50 mg, flurazepam 30 mg, and lormetazepam 2 mg) impaired driving the morning after dosing. In studies using driving simulators, diazepam (5–15 mg) produced deficits in lane position, speed maintenance, and emergency decisions. Diazepam (15 mg) impaired performance on a clinical test for drunkenness, which comprised 13 tests assessing motor, vestibular, mental, and behavioral functioning. Studies have also documented the effects of benzodiazepines in tracking tests in which subjects attempt to maintain a cursor within a moving target. Some tracking tests have been considered laboratory tests of driving ability because the device used to control the cursor was a steering wheel. Tracking performance is uniformly impaired by acute, therapeutic doses of numerous benzodiazepines.

Numerous studies have reported that acute doses of various benzodiazepines slow response time in simple or choice visual reaction time tests and impair attentional performance on the DSST and the Stroop test, a measure of selective attention. Benzodiazepines have also been shown to impair divided attention tests that require splitting attention between a central tracking task and responding to stimuli in the peripheral visual field. The impairment caused by benzodiazepines in tests of sustained attention or vigilance is not secondary to sedation, but rather a direct effect on perceptual sensitivity, resulting in decreased hits and increased response time in detecting stimulus targets.

Cannabis (marijuana)

Marijuana consists of the crushed leaves and stems of the plant *Cannabis sativa*. In the United States, marijuana is typically smoked, although in various parts of the world other preparations of the cannabis plant are eaten, or fumes from the ignited plant material are inhaled. The primary psychoactive ingredient of marijuana is Δ^9-tetrahydrocannabinol (THC), which currently averages 4.6% in commercial grade marijuana, but can range up to 25% in certain areas of the United States. Marijuana readily disrupts performance in complex tasks requiring continuous monitoring and the ability to shift attention rapidly between various stimuli. These same abilities are required when driving a car.

Marijuana has been shown to impair performance in driving simulators and laboratory tests that model various components of driving, such as reaction time, coordination, and tracking. Standardized field sobriety tests used by law enforcement officials to determine whether a person can drive safely were also shown to be impaired by marijuana. Several studies conducted in the 1970s indicated that marijuana impaired on-road driving abilities. A more recent comprehensive study of on-road driving found that marijuana moderately increased lateral movement of the vehicle within the driving lane on a highway.

Validity of Field Assessment Techniques

Background

The drug evaluation and classification (DEC) program was developed by the Los Angeles Police Department during the 1970s because of a need to identify and document whether a driver was impaired due to recent drug ingestion. The DEC program consists of a standardized evaluation conducted by a trained police officer (drug recognition examiner, DRE) and the toxicological analysis of a biological specimen. The evaluation involves a breath alcohol test, examination of the suspect's appearance, behavior, eyes, field sobriety tests (FSTs), vital signs, and questioning of the suspect. From the evaluation, the DRE concludes: (1) whether the suspect is behaviorally impaired such that he or she is unable to operate a motor vehicle safely; (2) whether the impairment is drug-related; and (3) the drug class(es) likely to be causing the impairment. The toxicological analysis either confirms or refutes the DRE's drug class opinion.

Several field studies have indicated that DREs' opinions were confirmed by toxicological analysis in 74–92% of cases when DREs concluded suspects were impaired. These studies attest to the validity of the DEC program as a measurement of drug-induced behavioral impairment in the field. However, the validity of the DEC evaluation has not been rigorously examined under controlled laboratory conditions.

Experimental studies

Recently, one of the authors (SJH) conducted two placebo-controlled, laboratory studies to examine the validity of the individual variables of the DEC evaluation as predictors of drug intake and to determine the accuracy of DREs in detecting whether subjects had been dosed with various drugs. In the first study, placebo and two active doses of ethanol, cocaine, and marijuana were administered to research volunteers before being evaluated by DREs.

Results showed that 17–28 variables of the DEC evaluation predicted the presence or absence of each of the three drugs with a high degree of sensitivity and specificity and low rates of false-positive and false-negative errors. In contrast, DREs' predictive accuracy of drug-induced impairment was not highly consistent with toxicological data. Of the 81 cases in which DREs predicted impairment, toxicology was positive for any drug(s) in 75 cases (92.6%). DREs' predictions were consistent with toxicology in 41 cases (50.6%), including 9 cases in which the DRE concluded impairment was due to ethanol alone. Because the DRE's breath test provided a priori confirmation of ethanol, an ethanol-only prediction was guaranteed to be consistent. Excluding those 9 cases resulted in 72 predictions that named a non-ethanol drug class. DREs' predictions were consistent with toxicology in 44.4% of cases.

In the second study, research volunteers were administered placebo and two active doses of alprazolam, *d*-amphetamine, codeine, or marijuana. The ability of the DEC evaluation to predict the intake of the four drugs or placebo was optimal when using 2–7 variables from the evaluation. DREs' decisions of impairment were consistent with the administration of any active drug in 76% of cases, and their drug class decisions were consistent with toxicology in 32% of cases.

Of particular relevance to the issue of drugs and driving safety is the part of the DEC evaluation in which four standardized FSTs are evaluated. The FSTs were Romberg balance (RB), walk and turn (WT), one-leg stand (OLS), and finger-to-nose (FN). The RB assessed body sway and tremor while subjects stood for 30 s with feet together, arms at sides, head tilted back, and eyes closed. The WT test required subjects to take nine heel-to-toe steps along a straight line marked on the floor, turn, and return with nine heel-to-toe steps. The OLS assessed balance by having subjects stand on one leg, with the other leg elevated in a stiff-legged manner 15 cm off the floor for 30 s. Subjects were given a brief rest between right and left leg testing. In the FN test, subjects stood as in the RB and brought the tip of the index finger of the left or right hand (as instructed) directly to the tip of the nose.

The broadest range of impairment was produced by alprazolam, which significantly increased body sway in the RB and disrupted balance in the WT and OLS. Marijuana impaired balance in the OLS, increased errors in the FN, and produced minor impairment in the WT. The relatively low doses of ethanol used in the study (see below) increased body sway in the RB and impaired coordination in the FN. Codeine increased errors in the OLS and FN. In the doses tested, neither cocaine nor *d*-amphetamine impaired performance on any of the tests; in fact, *d*-amphetamine decreased body sway in the OLS.

Plasma drug concentration

During both studies, repeated blood samples were taken to determine the relation between plasma drug concentration and impaired performance. In the first study, plasma concentration data are reported from the blood sample obtained halfway through the DEC evaluation. For ethanol, this was about 30 min after drinking ended, and ethanol (0.28 and $0.52\,g\,kg^{-1}$) produced mean $\pm$ standard error (SE) plasma concentrations of 24.3 ± 2.2 and $54.4 \pm 6.0\,mg\,dl^{-1}$, respectively. Cocaine (48 and $96\,mg\,70\,kg^{-1}$) produced mean plasma concentrations of 74.7 ± 7.2 and $180.5 \pm 17.1\,ng\,ml^{-1}$, respectively.

Benzoylecgonine concentrations were 95.4 ± 27.7 and $210.7 \pm 47.3\,ng\,ml^{-1}$, and ecgonine methylester levels were 10.8 ± 3.0 and $26.1 \pm 6.3\,ng\,ml^{-1}$ for low and high dose cocaine, respectively. Marijuana produced peak plasma concentrations immediately after smoking, which had declined at time of DEC evaluation to 15.4 ± 3.0 and $28.2 \pm 4.2\,ng\,ml^{-1}$ for 1.75% and 3.55% THC, respectively.

In the second study, plasma drug concentrations are reported for blood samples obtained 20 min before and immediately after the DEC evaluation. These blood samples corresponded to the following post-drug times: alprazolam, 60 and 105 min; *d*-amphetamine, 120 and 145 min; codeine, 60 and 105 min; and marijuana, 2 and 45 min. Mean $\pm$ SE plasma concentrations of alprazolam at 60 min postdrug were 12.9 ± 1.7 and $25.8 \pm 2.9\,ng\,ml^{-1}$ and at 105 min postdrug were 15.5 ± 1.5 and $31.3 \pm 2.0\,ng\,ml^{-1}$ for the 1 and 2 mg doses, respectively. *d*-Amphetamine produced mean $\pm$ SE plasma concentrations of 21.0 ± 2.5 and $42.8 \pm 4.2\,ng\,ml^{-1}$ at 120 min postdrug and 23.8 ± 1.6 and $46.6 \pm 3.4\,ng\,ml^{-1}$ at 145 min postdrug for the 12.5 and 25 mg doses, respectively. Codeine produced mean $\pm$ SE peak plasma concentrations at 60 min postdrug of 120.8 ± 14.8 and $265.8 \pm 28.8\,ng\,ml^{-1}$ for the 60

and 120 mg doses, respectively. At 105 min postdrug, plasma codeine concentrations were 114.1 ± 9.1 and $229.6 \pm 17.8\,\mathrm{ng\,ml^{-1}}$ for the low and high doses, respectively. Marijuana produced mean $\pm$ SE peak plasma THC concentrations 2 min after smoking of 28.1 ± 3.6 and $61.2 \pm 9.2\,\mathrm{ng\,ml^{-1}}$ for low and high doses, respectively. By 45 min postsmoking, THC concentrations had declined to 5.4 ± 0.7 and 9.9 ± 1.5 for low and high doses, respectively. Plasma THC levels of both doses were less than $1\,\mathrm{ng\,ml^{-1}}$ at 10 h postsmoking.

Conclusion

The validity of the DEC evaluation was examined by developing mathematical models based on discriminant functions that identified which subsets of variables best predicted whether subjects were dosed with placebo or each active drug. For all drugs except codeine, the subset of variables predicted the presence or absence of drug with a moderate to high degree of sensitivity and specificity. For codeine, sensitivity was low, and false-negative rates were extremely high. A secondary goal of these studies was to determine the accuracy of the DREs' evaluations in detecting whether an individual had been dosed with active drug. In the first study, non-ethanol drug class decisions were consistent with toxicology in 44% of cases. In the second study, DREs' drug class decisions were consistent with the administration of any active drug in 76% of cases, but consistent with toxicology in only 32% of cases. Thus, it would appear that DREs are able to detect drug-induced impairment in general, but have difficulty discriminating between various drugs.

These data clearly indicate that the variables of the DEC evaluation alone did not permit DREs to predict impairment and drug intake with the accuracy observed in field studies. There were several differences between the controlled laboratory conditions of this study and the field conditions under which DREs normally conduct the DEC evaluation that may have contributed to this outcome, such as the abridged form of the DEC evaluation, lack of meaningful drug-related cues (e.g. erratic driving, marijuana odor, drug paraphernalia), inability to interview subjects concerning drug use, and the possibility that greater drug doses encountered in the field would result in clearer behavioral signs of impairment. These findings suggest that predictions of impairment and drug use may be improved if DREs focused on a subset of variables associated with each drug class, rather than attempting to synthesize the entire DEC evaluation.

Postmortem Issues

Ethanol blood concentrations above a certain specified threshold are, by law, proof of intoxication. In theory, the same thresholds apply, even if the driver is dead. The situation becomes much more confusing when other drugs are involved. In some jurisdictions, the presence of any drug (parent or metabolite) is deemed proof of impairment. The question then arises as to whether postmortem blood drug concentrations in any way reflect drug concentrations at the time of death, and whether impairment can be inferred from the results of postmortem drug measurement.

At a minimum, postmortem drug concentrations depend on:

- the amount of drug ingested;
- the time elapsed from the last dose until the time of death;
- the time elapsed from death until specimen collection;
- where in the body the specimen is collected;
- sexual and racial differences in drug metabolism;
- drug tolerance (decreasing drug effect at a constant dose);
- the drug's volume of distribution.

Drug tolerance is almost certainly the most important factor. Concentrations of opiates and stimulants in trauma victims, where drugs have nothing to do with the cause of death, often exceed drug concentrations in decedents where the drug is clearly the cause of death. Unfortunately, after death, tolerance cannot even be estimated, let alone measured.

Factors effecting postmortem drug concentrations

Dose If illicit drugs are involved, the total amount of drug ingested is almost never known. Drug dealers, known as 'body stuffers', often ingest their inventory when arrest seems imminent. Even if the total number of 'rocks' swallowed is known, the amount of cocaine is not. Depending on the skill of the drugmaker, the rocks may contain much more bicarbonate than cocaine. Similar considerations apply to heroin. According to the Drug Enforcement Agency, retail heroin sold in the United States is, on average, 70% pure. But, unless a sample is available for analysis, there is no way to determine the purity of the fatal dose.

Premortem interval Parent drugs and metabolites generally have different half-lives. In the appropriate setting this difference can be used to estimate pattern of use, if not the exact time when the drug was taken. Very low concentrations of cocaine, in the face of high concentrations of benzoylecgonine, suggest that

cocaine was taken many hours before death. In the case of heroin, measurable concentrations of 6-monacetylmorphine (MAM), a heroin metabolite with an extremely short half-life, suggests use just prior to death. But unless death is witnessed and sudden, and an autopsy performed immediately, firm conclusions are impossible. The conversion of cocaine to benzoylecgonine continues after death, and heroin users may remain comatose for hours before expiring, allowing time for all the heroin to be converted to 6-MAM and thence to morphine.

Time to specimen collection After death, bacteria escape from the gastrointestinal tract. Some of these bacteria produce alcohol, while others metabolize it. Alcohol concentrations of $150\,mg\,dl^{-1}$, or higher, have been measured in alcohol-free cadavers stored for 2 days at room temperature. In the case of cocaine, if the postmortem interval is long enough, all the cocaine will be converted to benzoylecgonine, and ultimately to ecgonine. The changes that occur with other illicit drugs are unpredictable. Concentrations of drugs stored in fat, such as phencyclidine and marijuana, can be expected to increase, but the rate of increase has never been determined.

Site dependency Blood drug concentrations measured at autopsy depend on where the sample is obtained. Methamphetamine concentrations in left heart blood are 2–3 times higher than concentrations in the right heart, but only a tenth as high as concentrations measured in blood from the pulmonary artery. Cocaine concentrations in the subclavian artery decrease after death, while those in the femoral artery rise. In general, the increases are greater in central sites (heart and liver), and least in peripheral locations, such as the femoral artery. In individual cases, the changes are completely unpredictable. This situation comes about because of a process known as postmortem redistribution. After death, drugs diffuse along a concentration gradient. The process is more likely to occur if a drug is highly bound to protein, and if it is sequestered in a major organ, such as a lung or the liver. Redistribution begins immediately after death and continues indefinitely, although the biggest changes occur within the first 24 h, with greatest concentration increases occurring at central sights such as the heart and liver.

Sexual and racial differences The forensic importance of genetic polymorphism has only recently been recognized. Equivalent doses of ethanol produced higher peak blood levels in American Indians than in American Whites, and some speculate this difference may account for higher rates of alcoholism and death among American Indians. In a similar fashion CYP2D1 (the animal equivalent of CYP2D6 in humans) polymorphism appears to explain differential mortality rates observed after methamphetamine treatment in rats and presumably humans. Opiate metabolism also varies from individual to individual. The enzymatic conversion of hydrocodone to hydromorphone is catalyzed by cytochrome P_{450} 2D6, an enzyme that is inactive in about 7% of Caucasians (slow metabolizers). Death and impairment in poor metabolizers is likely to occur after much smaller amounts of drug have been ingested than in individuals with a normal cytochrome system. In addition, sex hormones also affect drug concentrations and responses. Women metabolize cocaine differently than men, and the same woman will metabolize cocaine differently at different times in her menstrual cycle. The list of drugs affected by such differences is extensive and growing longer.

Estimating drug concentrations at the time of death

Volume of distribution The volumes of distribution for morphine, cocaine and methamphetamine are all over $3\,l\,kg^{-1}$. As a consequence, less than 2% of a given dose is to be found circulating in the blood at any given moment. Since only 2% is actually circulating in the blood, the release of only 1% more from deep tissue stores would be enough to double observed postmortem tissue concentrations. In controlled animal studies, levels of free morphine, for example, more than doubled during the first 24 h after death. The situation becomes even more complicated when metabolites are considered. The volume of distribution for morphine is more than 10 times that of either morphine-3-glucuronide or morphine-6-glucuronide. In contrast to free morphine, which diffuses into tissues throughout the body, almost all the conjugated morphine is to be found circulating in the blood. Any inference based on measurements of total morphine circulating in the blood may either wildly over- or underestimate the actual situation at the time of death.

Ethanol estimates Premortem ethanol measurements are, by statute, made with plasma, not whole blood. Postmortem alcohol measurements are made on whole blood, which contains 10–15% less water than serum or plasma (the red cells take up room). In practice, the difference may be much greater, because all tissues in the body lose water after death, including blood. As water is lost, the concentration of alcohol, changes. Still, provided the appropriate specimens are obtained at the time of autopsy, it is generally possible to distinguish antemortem intoxication from

postmortem alcohol production. One way to do this is by measuring vitreous alcohol. Vitreous fluid contains more water than blood, and a blood:vitreous ratio >0.95 is a good indication that death occurred before the blood and vitreous were in equilibrium, suggesting that blood alcohol concentrations (BACs) were rising at the time of death. Similarly, a urine alcohol concentration (UAC):BAC ratio <1.2 is good evidence that blood concentration was rising at the time of death. UAC:BAC ratios greater than 1.3 are consistent with a postabsorptive stage at the time of death. Ratios much greater than 1.3 suggest heavy consumption over a long period of time. The relationship between BAC and UAC is not accurate enough for forensic purposes, but BACs can be estimated by dividing autopsy UAC by 1.35.

Conclusions

Except for alcohol, where impairment is defined by statute, relating impairment to specific blood concentrations of any abused drug is, even in the living, problematic. More often than not, it is impossible. After death, the issues become much more complex. Given our current state of knowledge, inferring impairment from any postmortem drug measurements (except perhaps alcohol) is simply not possible.

See also: **Alcohol**: Blood. **Drugs of Abuse**: Postmortem Blood.

Further Reading

Chait LD and Pierri J (1992) Effects of smoked marijuana on human performance: a critical review. In: Murphy L and Bartke A (eds) *Marijuana/Cannabinoids: Neurobiology and Neurophysiology*, pp 387–423. Boca Raton, FL: CRC Press.

Hearn WL, Keran EE, Wei HA and Hime G (1991) Site-dependent postmortem changes in blood cocaine concentrations. *Journal of Forensic Sciences* **36**(3):673–684.

Heishman SJ (1998) Effects of abused drugs on human performance: laboratory assessment. In: Karch SB (ed.) *Drug Abuse Handbook*, pp 206–235. Boca Raton, FL: CRC Press.

Heishman SJ, Singleton EG and Crouch DJ (1998) Laboratory validation study of drug evaluation and classification program: Alprazolam *d*-amphetamine, codeine, and marijuana. *Journal of Analytical Toxicology* **22**:503–514.

Jones G (1998) Interpretation of postmortem drug levels. In: Karch SB (ed.) *Drug Abuse Handbook*, pp 970–987. Boca Raton, FL: CRC Press.

Karch SB (1996) *The Pathology of Drug Abuse*, 2nd edn. Boca Raton, FL: CRC Press.

Koelega HS (1993) Stimulant drugs and vigilance performance: a review. *Psychopharmacology* **111**:1–16.

Lukas SE, Sholar M and Lundahl LH *et al.* (1996) Sex differences in plasma cocaine levels and subjective effects after acute cocaine administration in human volunteers. *Psychopharmacology (Berl)* **125**(4):346–354.

O'Hanlon JF, Vermeeren A, Uiterwijk M, van Veggel L and Swijgman HF (1995) Anxiolytics' effects on the actual driving performance of patients and healthy volunteers in a standardized test. *Neuropsychobiology* **31**:81–88.

Pounder DJ and Jones GR (1990) Post-mortem drug redistribution: a toxicological nightmare. *Forensic Science International* **45**(3):253–263.

Prouty RW and Anderson WH (1990) The forensic science implications of site and temporal influences on postmortem blood-drug concentrations. *Journal of Forensic Sciences* **35**(2):243–270.

Ramaekers JG (1998) Behavioral toxicity of medicinal drugs. *Drug Safety* **18**:189–208.

Robbe HWJ (1994) *Influence of Marijuana on Driving*. Maastricht: University of Limburg Press.

Segal B and Duffy LK (1992) Ethanol elimination among different racial groups. *Alcohol* **9**(3):213–217.

Hair

P Kintz, Institut de Médecine Légale, Strasbourg, France

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0492

Introduction

It is generally accepted that chemical testing of biological fluids is the most objective means of diagnosis of drug use. The presence of a drug analyte in a biological specimen can be used to document exposure. The standard in drug testing is the immunoassay screen, followed by gas chromatographic–mass spectrometric confirmation conducted on a urine sample. In recent years, remarkable advances in sensitive analytical techniques have enabled the analysis of drugs in unconventional biological specimens, such as saliva, sweat, meconium and hair. The advantages of these samples over traditional media, like urine and blood, are obvious: collection is almost noninvasive, is relatively easy to perform, and in forensic situations may be achieved under close supervision of law enforcement officers to prevent adulteration or substitution. Moreover, the window of drug detection is dramatically extended to weeks, months or even

years. The aim of this article is to document the usefulness of these alternatives in forensic situations.

Hair testing for drugs of abuse in humans was first demonstrated in 1979 in the United States, and was rapidly followed by some German results. Since then, more then 300 papers have been published, most of them being devoted to analytical procedures. After an initial period during which drugs were analyzed using immunological tests, the standard is now to use gas chromatography coupled with mass spectrometry. During the early stages of hair testing, opiates and cocaine were the predominant analytes, followed by amphetamine derivatives. Only recently have cannabis, benzodiazepines and, very recently, doping agents been evaluated.

Hair analysis was initially developed to document forensic cases; however, today, numerous applications have been described in clinical, occupational and sporting situations.

Incorporation of Drugs in Hair

This is one of the major points of disagreement among scientists actively involved with hair analysis. The time at which a drug appears in hair after administration is highly variable. According to several authors, this delay can be some hours or even days. Based on these differences, a complex model has been proposed to account for drug incorporation. Both sweat and sebum have been suggested to complement blood in this process but the exact mechanism is still under discussion. It has been demonstrated that, for the same administered dose, black hair incorporates more drug than blond hair, clearly indicating the influence of melanin. For some authors, these findings suggest a racial element in hair composition. Cosmetic treatments, like bleaching or waving, affect the drug content, producing a 50–80% reduction in the original concentration.

In almost all cases, the major compound detected in hair is the parent drug, much more so than its metabolites. For example, cocaine is present at concentrations 5–10 times greater than benzoylecgonine, and 10–30 times greater than ecgonine methylester, although, in blood, both metabolites are found at higher concentrations than cocaine. The very short half-life of heroin and 6-acetylmorphine makes their detection quite impossible in blood, but these two compounds are found in larger amounts than morphine in hair. This is also the case for sweat, and thus confirms the implication of sweat in the incorporation of drug in hair.

Environmental contamination has also been proposed as a potential risk of incorporation, leading to false positives. Drugs that are smoked, like cannabis, crack or heroin, are of concern, and it is therefore necessary to include a decontamination step to eliminate false-positive findings. Various procedures have been described in the literature: these involve the use of organic solvents, aqueous solutions or a sequence of solvent and buffer. To minimize the influence of external contamination, several authors have proposed various appraches including an analysis of the wash solution, or to carry a kinetic of the wash ratios. Others have proposed the identification of specific or unique metabolites, such as norcocaine or cocaethylene, a compound that is formed when concomitant cocaine and ethanol are used. The detection of specific markers is not easy, as their concentrations in hair are generally low. Therefore, some authors have proposed the use of drug ratios, like morphine-codeine, to document heroin abuse, or benzoylecgoninecocaine greater than 0.05, to document cocaine abuse. Finally, positive cutoffs have been published to insure international uniformity (**Table 1**).

After collection, the hair specimen is best stored at ambient temperature. Once incorporated in hair, drugs are very stable. Cocaine and benzoylecgonine have been detected in hair from Peruvian mummies, clearly demonstrating 500 years of stability.

Hair Collection and Analysis

Strands of hair (about 60 to 80) are cut as close as possible to the skin, in the posterior vertex region, dried and stored in tubes at room temperature. The root-to-end direction must be indicated. In the case of very short hair, pubic hair can also be collected.

Typical hair preparation involves the following steps:

- decontamination of the strand in organic solvent or buffer;
- pulverization of 100 mg in a ball-mill;
- hydrolysis of a 50 mg sample in acid or alkaline buffer;
- purification by solid-phase or liquid–liquid extraction;
- derivatization;
- analysis with gas chromatography coupled with mass spectrometry (GC-MS).

Table 1 Proposed positive cutoffs

Analytes	*Cutoff (ng mg^{-1})*
Heroin	0.5 for 6-acetylmorphine
Cocaine	0.5 for cocaine
Amphetamine, MDMA	0.5 for both drugs
Cannabis	Not decided

Care is necessary to prevent the conversion of cocaine to ecgonine, or 6-acetylmorphine to morphine, in alkaline solution. Differences in efficiency between enzymatic and acid hydrolysis are not statistically significant.

A critical element in the acceptance of hair analysis for detection of drugs of abuse is laboratory performance. Laboratories must be able to demonstrate that they can accurately determine what drugs are present in unknown hair samples and at what concentrations. Several international interlaboratory comparisons of qualitative and quantitative determinations of drugs have been organized in the United States (National Institute of Standards and Technologies, Gaithersburg, MD) and Europe (Society of Hair Testing, Strasbourg, France). Interlaboratory comparisons of hair analysis have been published for opiates, cocaine, cannabis and amphetamines. In most cases, GC-MS was used for the analyses. However, no one extraction method could be identified as being superior to others.

In 1999, the following compounds were reported to be detectable in hair:

drugs of abuse (opiates, cocaine, cannabis, amphetamines, methadone, phencyclidine, narcotics);
pharmaceuticals (barbiturates, antidepressants, benzodiazepines, neuroleptics, etc.); nicotine; doping agents (β-adrenergic drugs, anabolic steroids and corticosteroids);
pesticides.

Measured concentrations are expressed in picograms or nanograms per microgram.

Detection of Drugs of Abuse in Hair

Opiates

Three methods of screening for opiates, cocaine, cannabinoids and amphetamine, including its derivatives, dominate in the literature; these are briefly described in **Table 2**. Liquid–liquid extraction after HCl hydrolysis and solid-phase extraction after enzymatic hydrolysis with β-glucuronidase/sulfatase lead to similar results, both with the disadvantage that heroin and 6-O-acetylmorphine (MAM) might be hydrolyzed to morphine. The methanol method is undoubtedly the simplest, with high sensitivity for heroin and cocaine but poor sensitivity for their metabolites, morphine and benzoylecgonine; and high sensitivity for Δ^9-tetrahydrocannabinol (THC) but no sensitivity for THC-COOH. In 1995, it was confirmed by systematic extraction studies that

Table 2 Screening procedures for the detection of illegal drugs in hair

	Kauert (1996)	*Kintz (1995)*	*Moeller (1993)*
Analytes	Heroin, 6-MAM, dihydrocodeine, codeine, methadone, THC, cocaine, amphetamine, MDMA, MDEA, MDA	Heroin, 6-MAM, dihydrocodeine, codeine, methadone, THC, cocaine, amphetamine, MDMA, MDEA, MDA	Heroin, 6-MAM, dihydrocodeine, codeine, methadone, THC, cocaine, amphetamine, MDMA, MDEA, MDA
Decontamination step	Ultrasonic 5 min each 5 ml H_2O 5 ml acetone 5 ml petrolether	5 ml Cl_2CH_2 (2 × 5 min)	20 ml H_2O (2 ×) 20 ml acetone
Homogenization	100 mg hair cut into small sections in a 30 ml vial	Ball-mill	Ball-mill
Extraction	4 ml methanol ultrasonic 5 h at 50°C	50 mg powdered hair, 1 ml 0.1 N HCl, 16 h at 56°C	20–30 mg powdered hair, 2 ml acetate buffer + β-glucuronidase/ arylsulfatase, 90 min at 40°C
Clean-up	None	$(NH_4)_2HPO_4$; extraction 10 ml CHCl3/2-propanol/*n*-heptane (50:17:33); organic phase purified with 0.2 N HCl; HCl phase to pH 8.4; re-extraction with $CHCl_3$	$NaHCO_3$; SPE (C18), elution with 2 ml acetone/CH_2Cl_2 (3:1)
Derivatization	Propionic acid anhydride	40 µl BSTFA/1% TMCS; 20 min at 70 °C	1000 µl PFPA/75 µl PF-*n*-propanol; 30 min at 60°C; N_2 at 60°C; 50 µl ethylacetate

MDEA: methylenedioxyethylamphetamine; SPE: solid phase extraction; PFPA pentafluoropropionic anhydride; PF: pentafluoropropanol; BSTFA: N,O-bis(trimethylsilyl)trifluoroacetamide; TMCS: trimethylchlorosilane.

methanol and water had the best extraction capability for opiates, but using hydrophobic solvents like dioxane and acetonitrile, a low extraction rate was found. With toluene, almost no extraction occurred. The range of positive results using these procedures is listed in **Table 3**. Pubic hair showed higher drug levels than scalp hair. This can be due to the slightly lower growth rate of pubic hair than scalp hair but, additionally, pubic and scalp hair have totally different telogen:anagen ratios and concentrations cannot be directly compared. Regarding individual growth rate and the problem of telogen:anagen ratios, dose–concentration relation studies should only be performed with hair samples grown from the shaved skin before drug administration and under control of the growth speed of the hair.

In hair specimens of 20 subjects receiving intravenous heroin hydrochloride, no correlation between the doses administered and the concentrations of total opiates in hair was observed. However, when considering a single analyte, it was noted that the correlation coefficient seemed to be linked to its plasma half-life. A weak correlation coefficient corresponds to a drug with a short plasma half-life, and the correlation coefficient increases when plasma half-life increases, from heroin, 6-acetylmorphine to morphine.

The so-called poppy seed problem could by solved by examining hair for morphine after poppy seed ingestion, as morphine is not detected in hair after consumption of seeds, or at least only in traces.

Cocaine

The fact that the parent drug is found in higher concentrations in the hair of drug users has been well known since 1991. Typical concentration ranges are listed in **Table 4**.

Contrary to the case in heroin abuse, cocaine consumption can be detected by measurable metabolites which cannot be caused by cocaine contamination, like norcocaine or cocaethylene. The determination of the pyrolysis product of cocaine, the anhydroecgonine methyl ester (AEME), can be helpful in distinguishing cocaine and crack users.

The literature and the scientific meetings concerning cocaine are dominated by discussion as to whether decontamination procedures can remove external contamination completely, and whether a racial element exists. This is important when hair analysis is used as 'stand-alone' evidence for workplace testing.

Table 3 Published ranges of 6-*O*-acetylmorphine concentration in the hair of heroin users

Authors	*Concentration* ($ng\,mg^{-1}$)
Kauert (1996)	0.03–79.8
Kintz (1995)	0–84.3
Moeller (1993)	2.0–74
Pepin (1997)	0.3–131.2

Table 4 Published ranges of cocaine concentration in the hair of drug users

Authors	*Concentration* ($ng\,mg^{-1}$)
Kauert (1996)	0.04–129.7
Kintz (1995)	0.4–78.4
Moeller (1993)	0.3–127.0
Pepin (1997)	0.89–242.0

An important study with controlled doses of cocaine-d_5 was published in 1996. The deuterium-labeled cocaine was administered intravenously and/or intranasally in doses of 0.6–4.2 $mg\,kg^{-1}$ under controlled conditions. A single dose could be detected for 2–6 months; the minimum detectable dose appeared to be between 22 and 35 mg; but within the range of doses used in the study, the hair test did not provide an accurate record of the amount, time or duration of drug use.

Cocaine, benzoylecgonine and ecgonine methylester have also been found in the mummified bodies of ancient Peruvian coca-leaf chewers. In contrast to today's cocaine users, the cocaine:benzoylecgonine ratio was less than 1.

Cannabis

In 1996, the first results on levels of cannabis in hair measured by using GC-MS were reported, simultaneously with the determination in the same run of THC and its major metabolite THC-COOH. The measured concentrations were low, particularly in comparison with other drugs. Some authors suggested the use of negative chemical ionization to target the drugs, or the application of tandem mass spectrometry. More recently, a simpler method was proposed, based on the simultaneous identification of cannabinol, cannabidiol and THC. This procedure appears to be a screening method that is rapid and economical and does not require derivatization prior to analysis. As THC, cannabinol and cannabidiol are present in smoke, to avoid potential external contamination the endogenous metabolite THC-COOH should be secondarily tested to confirm drug use.

As shown in **Table 5**, the concentrations measured are very low, particularly for THC-COOH, which is seldom identified. To date, there is no consensus on cutoff values for cannabis. An international debate must be held to discuss the differences noted between

Table 5 Reported concentrations for cannabis in hair

Authors	*Compound*	*Number of positives*	*Concentration (ng/mg)*
Cairns (1995)	THC-COOH	+ 3000	(0.0007)
Jurado (1995)	THC	298	0.06–7.63 (0.97)
	THC-COOH	298	0.06–3.87 (0.50)
Kauert (1996)	THC	104	0.009–16.70 (1.501)
Kintz (1995)	THC	89	0.10–3.39 (0.64)
	Cannabidiol	306	0.03–3.00 (0.51)
	Cannabinol	268	0.01–1.07 (0.16)
	THC-COOH	267	0.05–0.39 (0.10)
Moeller (1993)	THC	10	0.4–6.2 (2.0)
	THC-COOH	2	1.7–5.0 (3.3)
Wilkins (1995)	THC	8	0.03–1.1

Values in parentheses are mean concentrations. MDEA: methylenedioxyethylamphetamine; MBDB: N-methyl-1-(1,3-benzodioxol-5-yl)-2-butanamine; BDB: 1-(1,3-benzodioxol-5-yl)-2-butanamine.

American laboratories, which reported THC-COOH in the low picogram per milligram range, and some European laboratories, which gave concentrations in the low nanogram per milligram range, as is obvious from the measured concentrations shown.

Amphetamine derivatives

Almost all of the literature dealing with amphetamines in hair has been written by Japanese researchers. In most cases, amphetamine and methamphtetamine have been the target drugs. More recently, particular attention has been focused on methylenedioxyamphetamine (MDA) derivatives, like methylenedioxymethamphetamine (MDMA). Most techniques published used acid or alkaline hydrolysis, or a combination of hydrochloric acid and methanol, followed by a purification step (liquid–liquid extraction or solid-phase extraction) and derivatization.

When comparing four different procedures for amphetamine, MDA and MDMA (methanol sonication, acid hydrolysis, alkaline hydrolysis and enzymatic hydrolysis) it was demonstrated that best recovery rates were observed after alkaline hydrolysis; however, it was not possible to determine which method performed best, based on recovery rate, precision and practicability. Lower concentrations were observed after methanol sonication, together with dirty chromatograms.

It must be pointed out that, since the first identification of MDMA in human hair in 1992, this compound, particularly in Europe, is one of the most frequently identified and must therefore be included in all screening procedures.

Typical findings for amphetamine derivatives are given in **Table 6**.

Although there are still controversies surrounding the interpretation of results, particularly concerning external contamination, cosmetic treatments, ethnic bias or drug incorporation, pure analytical work in hair analysis has more or less reached a plateau, having solved almost all the analytical problems. Conferences on hair analysis in Genoa, Strasbourg, Tampa and Abu Dhabi, between 1992 and 1996, indicate the increasing role of this method for the investigation of drug abuse.

Table 6 Analytical parameters and results for a general screening procedure for amphetamine derivatives

Compound	*Ions monitored (m/z)*	*Linearity (r)*	*Precision (at 2 ng mg^{-1}, %)*	*Concentration (ng mg^{-1})*
Amphetamine	91, 118, 240	0.998	6.9	2.3–20.6 (*n* = 5)
Methamphetamine	169, 210, 254	0.995	8.4	
MDA	135, 240, 375	0.994	9.1	0.4–8.0 (*n* = 13)
MDMA	210, 254, 389	0.996	10.2	0.3–42.7 (*n* = 14)
MDEA	240, 268, 403	0.997	13.0	0.6–69.3 (*n* = 6)
MBDB	176, 268, 403	0.994	8.7	1.41–3.09 (*n* = 2)
BDB	135, 176, 389	0.996	9.4	0.21 (*n* = 1)

Values in parentheses are numbers of positive cases.

Applications

In the case of segmental analysis, to evaluate the pattern of drug abuse for example, proximal and distal portions of hair must be identified. Given the variation in hair growth rates, generally 1.0–1.3 cm per month, results from a multisectional analysis should not be used to determine a precise period of drug exposure. The further away from the hair root, the more cautious the interpretation of quantitative findings of the individual hair sections must be.

Table 7 lists the major characteristics of both urine and hair analyses.The major practical advantage of hair testing compared with urine testing for drugs is its larger surveillance window: weeks to months in hair, depending on the length of the hair shaft, versus 2–4 days in urine for most xenobiotics, except cannabis. In fact, for practical purposes, the two tests complement each other. Urinalysis provides short-term information on an individual's drug use, whereas long-term histories are accessible through hair analysis. While analysis of urine specimens cannot distinguish between chronic use or single exposure, hair analysis makes this distinction. Its greatest use, however, may be in identifying false negatives, as neither abstaining from a drug for a few days nor trying to 'beat the test' by diluting urine will alter the concentration in hair. Urine does not indicate the frequency of drug intake in subjects, who might deliberately abstain for several days before screening.

It is always possible to obtain a fresh, identical hair sample if there is any claim of a specimen mix-up or breach in the chain of custody. This makes hair analysis essentially fail-safe, in contrast to urinalysis, as an identical urine specimen cannot be obtained at a later date. Clearly, hair analysis can thus function as a 'safety net' for urinalysis.

Numerous forensic applications have been described in the literature where hair analysis was used to document the case: differentiation between a drug dealer and a drug consumer, chronic poisoning, crime under the influence of a drug, child sedation, child abuse, doubtful death, child custody, abuse of drugs in jail, body identification, survey of drug addicts, chemical submission, obtaining a driving license and doping control. It appears that the value of hair analysis for the identification of drug users is steadily gaining recognition. This can be seen from its growing use in preemployment screening, in forensic sciences and in clinical applications. Hair analysis may be a useful adjunct to conventional drug testing in toxicology. Methods for evading urinalysis do not affect hair analysis. Specimens can be more easily obtained with less embarrassment, and hair can provide a more accurate history of drug use. Costs are too high for routine use but the generated data are extremely helpful in documenting positive cases.

See also: **Drugs of Abuse**: Analysis. **Toxicology**: Overview; Methods of Analysis – Antemortem; Methods of Analysis – Postmortem; Interpretation of Results. **Hair**: Hair Transfer, Persistence and Recovery; Comparison: Other.

Further Reading

Baumgartner WA, Hill VA and Blahd WH (1989) Hair analysis for drugs of abuse. *Journal of Forensic Sciences* **34**:1433–1453.

Cairns T, Kippenberger D and Gordon AM (1997) Hair analysis for detection of drugs of abuse. In: Wong SHY and Sunshine I (eds) *Analytical Therapeutic Drug Monitoring and Toxicology*, pp. 237–251. Raton, FL: CRC Press.

de Zeeuw R, Al Hosani I, Al Munthiri S and Maqbool A (eds) (1995) Hair analysis in forensic toxicology. *Proceedings of the 1995 International Conference and Workshop*, pp. 1–503. Abu Dhabi: Abu Dhabi Police.

Forensic Science International (1993) Special issue: Hair analysis as a diagnostic tool for drugs of abuse investigation, edited by Tagliaro F, Chiarotti N, and Deyl Z.

Table 7 Comparison between urine and hair for testing drugs of abuse

Parameter	*Urine*	*Hair*
Drugs	All	All
Major compound	Metabolites	Parent drug
Detection period	2–4 days, except cannabis	Weeks, months
Type of measure	Incremental	Cumulative
Screening	Yes	Yes
Invasiveness	High	Low
Storage	$-20\ ^{\circ}C$	Ambient temperature
Risk of false negative	High	Low
Risk of false positive	Low	Undetermined
Risk of adulteration	High	Low
Control material	Yes	Yes

Forensic Science International (1995) Special issue: Hair analysis II. A workshop held at the XVIth Congress of the International Academy of Legal Medicine and Social Medicine, Strasburg, France, 31st May–2 June, 1994, edited by Kintz P.
Forensic Science International Special issue: Proceedings of the 1st European meeting on hair analysis. Clinical, occupational and forensic application, Genova, Italy, June 17–19, 1999 edited by Kuitz P.
Forensic Science International (2000) Special issue: Selected papers from the 2nd international meeting of the Society of Hair Testing edited by Kintz P, **107**:1–402.
Harkey MR and Henderson GL (1987) Hair analysis for drugs of abuse. In: Baselt RC (ed.) *Advances in Analytical Toxicology*, pp. 298–329. Chicago: Year Book.
Inoue T, Seta S and Goldberger B (1995) Analysis of drugs in unconventional samples. In: Liu R and Goldberger B (eds) *Handbook of Workplace Drug Testing*, pp. 131–158. Washington, DC: AACC Press.
Kintz P (ed) (1993) Forensic scientific examination of human hair: detection of drugs. In: Wecht C (ed.) *Forensic Sciences*, 49–32. New York: Matthew Bender.
Kintz P (ed) (1996) *Drug Testing in Hair*. Boca Raton, FL: CRC Press.
Kintz P (1996) Drug testing in addicts: a comparison between urine, sweat and hair. *Therapeutic Drug Monitoring*. **18**:450–455.
Kintz P (1998) Hair testing and doping control in sport. *Toxicology letters* **102–103**:109–113. New York: Matthew Bender.
Kintz P (2000) Quelle place pour les cheveux dans la lutte contre le dopage? *Annales de Toxicologie Analytique* **12**:49–55.
Sachs H and Kintz P (1998) Testing for drugs in hair. Critical review of chromatographic procedures since 1992. *Journal of Chromatography B* **713**:147–161.

Postmortem Blood

M D Osselton, Forensic Science Service, Reading, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0489

Introduction

Postmortem blood samples may be used by forensic toxicologists as an aid to determining whether a deceased person might have been affected by a drug or poison at the time of death or whether the drug or poison detected could have been directly responsible for death. Recent research has shown, however, that the use of postmortem blood drug concentrations for interpretative purposes must be undertaken circumspectly, and their comparison against published data collections should be undertaken with extreme caution and only after careful consideration of all of the factors that could potentially affect the analytical results.

Sample Collection

Following a sudden or unexpected death the forensic toxicologist is often requested to undertake toxicological analysis to determine whether drugs or other poisons may have contributed to or caused death. The common circumstances where analysis may be requested are:

- after a suspected drug overdose;
- after a suspected poisoning;
- in accident investigation – to determine whether drugs may have been a contributory factor to an accident;
- after a death linked with a possible criminal offence (e.g. assault/murder).

In order to investigate whether drugs and/or poisons may have contributed to the cause of death, or whether the deceased may have been under the influence of drugs at the time of death, blood, among other tissue samples, can provide valuable information. A general outline concerning blood as a tool for toxicological analysis is presented elsewhere in this work and the reader is referred to the section dealing with antemortem blood samples for a general discussion on the matrix of blood. There are, however, a number of significant considerations that must be borne in mind when analysing postmortem tissues that are unique to these samples; a different approach to interpretation is warranted than that which is applied to blood samples obtained from living subjects.

In most instances it is possible to collect a range of tissues at postmortem examination and this should be the recommended procedure whenever possible. It is the duty of the toxicologist to ensure that the pathologist is provided with advice concerning the samples required and appropriate containers. Whenever possible, the laboratory should provide a sample collection kit containing suitable containers, labels and packaging to enable the pathologist to collect and submit a full range of samples. It is better to collect more samples than may be required at the first postmortem examination, rather than having to try to go back for more later.

The postmortem specimen collection kit should include, at a minimum:

- three 10 ml screw-cap containers and containing oxalate for peripheral blood samples;
- one 30 ml universal container with screw cap for a urine sample;
- two 5 ml fluoride oxalate vials with septum caps for blood alcohol/volatiles analysis;
- one 5 ml fluoride oxalate vial for urine alcohol analysis;
- two large screw-cap containers or jars for the collection of liver and stomach contents;
- three 20 ml syringes with needles;
- set of tie-on labels (sufficient for all containers);
- laboratory submission/information form;
- secure transportation box;
- instruction sheet outlining collection requirements.

There is no universal test for a drug or poison. So, when the availability of tissue/samples is limited, the analyst should have all of the information available to assist in conserving valuable samples for the most appropriate analysis. Access to a full case history may not only help to save valuable samples but can also save considerable time and expense in carrying out unnecessary tests. The investigating officer should therefore be required to submit as much information about the case as possible, as this may influence the type and extent of analysis undertaken as well as the interpretation of analytical results. Submitting officers should be required to provide, at a minimum, the following information:

- name of deceased;
- age;
- occupation;
- race (Caucasian, Chinese, Afro Caribbean, etc.);
- date/time of discovery of body;
- date/time of death;
- date/time of postmortem;
- time between death and postmortem;
- full details/case history and circumstances surrounding/leading up to death;
- details of any signs or symptoms noted prior to death (e.g. diarrhoea, loss of weight, delirium, drunkenness, convulsions, hallucinations, etc);
- temperature of body surroundings/environment where body was found;
- medical history of deceased;
- availability to deceased or relatives of drugs/medicines/poisons;
- whether deceased was a known or suspected drug user;
- whether the body was embalmed prior to autopsy;
- if death occurred following admission to hospital, the survival period between admission and death;
- the treatment/drugs/resuscitation procedures carried out before death;
- if death occurred in hospital, details of any antemortem blood samples collected (if any of these are available they should be requested for analysis);
- whether the stomach was washed out on admission to hospital (if so, access to the stomach wash, if it is available, should be requested).

All too often a seemingly straightforward and uncomplicated death can turn into a serious case that will culminate in some form of criminal proceedings. It is far better to have all of the information at hand while planning how to manage the case than to have to revisit it in hindsight or at the request of the defense lawyer. The more information that is at hand before undertaking any work, the more reasoned the approach to analysis and interpretation will be. With the increasing availability of sensitive bench-top mass spectrometry to almost all laboratories in recent years, there has been a trend towards undertaking blood analysis alone in many cases, rather than performing analysis on a range of tissues. Historically, blood, stomach contents, liver and urine were all used to provide a composite picture in postmortem toxicology cases. This was partly necessitated because analytical methods were relatively insensitive and necessitated a plentiful supply of material for drug extraction. With the advent of gas chromatography and inexpensive bench-top mass spectrometry, toxicologists have been able to develop analytical methods that enable blood analyses to be undertaken for a wide range of drugs in small blood samples, often at concentrations significantly lower than in liver, stomach content and urine. The introduction of these new technologies led to the publication of many data compilations listing the concentrations of drugs in blood following the consumption of therapeutic quantities, in addition to those measured where patients had taken toxic or overdose quantities of drugs. The increasing availability of data relating to concentrations of drugs in blood led many toxicologists to direct their analysis towards blood only, particularly when this could be accompanied by minimizing the costs and time spent on undertaking multiorgan toxicology. The concept that whole blood analysis would provide an estimation of the concentration of a drug at the time of death is one that was held in wide belief and led to a widespread fall in laboratories undertaking multiorgan toxicology. During recent years our knowledge concerning what happens to drugs post mortem, between death and the time of sampling, has grown and led many forensic toxicologists to view with caution the way that analytical results are used. If blood samples could be collected within minutes of death, the philosophy that a postmortem blood sample provides a ‘snapshot’ of the blood

concentration of a drug at the time of death might be supportable. In reality, however, it is usually impossible to collect blood samples within minutes of death. In most cases a considerable amount of time elapses between the moment of death and postmortem blood sampling. In the majority of cases encountered by the forensic toxicologist, the time of death and the interval elapsing between death and sampling is a totally unknown entity.

Postmortem Change

The living person comprises millions of microscopic cells grouped together as tissues or organs, each with specific functions to perform. Each cell is a highly organized structure of chemical membranes, containing and controlling numerous complex chemical processes in a highly organized manner. The cells are bathed by a fluid, the interstitial fluid, that provides a link with the circulating blood, the function of which is to transport oxygen and essential substances around the body for the cells and to carry waste products and toxins away from the cells to specific sites in the body from where they may be eliminated (e.g. liver and kidneys). When death occurs, the mechanisms for controlling the chemical reactions within the cells cease to function. Failure of the heart to circulate blood deprives the body tissues of oxygen and other nutrients and leads to the accumulation of waste materials in and around the cells, resulting, within a short time, in cell death. Cell death is characterized by chemical disorganization, failure of cell metabolism and cell function, and eventually a disintegration of the cell structure. As the cell walls begin to lose integrity, leakage of the cell contents into the surrounding environment takes place and 'out of control' enzymes, formerly contained within the cells, start to destroy the cells and tissues in a self-destructive process known as autolytic decomposition.

After death, the temperature at which the body is stored can alter the rate of decomposition. In a body stored under refrigerated conditions the process of autolysis and decomposition is slowed down, whereas a body maintained at an elevated temperature will be likely to decompose at a faster rate. After death, the processes that combat and protect the body from microbial infection become ineffective, hence microbial action also contributes to the process of decomposition. As decomposition advances, drugs may be released from tissues and intracellular binding sites and diffuse from areas of higher concentration to areas of lower concentration. This offers an explanation as to why postmortem heart blood concentrations may be several orders of magnitude higher than blood collected from the femoral vein in the upper leg. Following the consumption of a drug overdose, the liver concentrations of drugs can become very high, as this organ is one of the principal organs in the body where drugs are metabolized in preparation for their elimination. Likewise, capillary-rich lung tissue containing lipoproteins may also accumulate high concentrations of drugs. After death, drugs diffuse from the lung, liver and other tissues where they are present in high concentrations into the pulmonary artery and vein and also into the vena cava. The diffusion process continues over time, resulting in significant variations in blood drug concentrations.

The observation that drugs are found in different concentrations at different sites in the body after death, was first recorded in 1960. This observation prompted the recommendation that, ideally, peripheral blood samples should be used for toxicological analyses. It was not, however, until the mid to late 1980s that toxicologists really began to realize the full significance of postmortem change. Several groups published a number of significant observations that have led toxicologists to express caution in interpreting postmortem blood drug concentrations. One group demonstrated that postmortem redistribution of basic drugs can occur rapidly after death and that attempts to resuscitate a corpse using cardiac massage, or even movement of the body from the site of death to the postmortem room, may influence drug distribution. As a result of their observations, the authors judged that it is unsafe to attempt to apply pharmacokinetic calculations, based on the analysis of a postmortem blood analysis, to deduce the quantity of drug taken. It is also unsafe to assume that blood drawn from a peripheral site is unaffected by postmortem redistribution. Pathologists or postmortem room technicians must be aware that simply withdrawing blood down the femoral vein may result in drawing blood down from the major vessels in the torso, such as the vena cava, where the process of diffusion from the liver or heart may be well advanced.

The ability to obtain a satisfactory peripheral blood sample at post mortem may be affected by the state of the body and blood vessels. Ideally, the femoral vein should be dissected and cut and the portion closest to the abdomen tied off with a ligature. Blood should then be collected from the leg portion of the vein, either using a syringe or by gently massaging the vein along the leg. Many pathologists prefer to collect blood samples from the vein in the upper arm rather than the leg. However, since the subclavian vein is much closer to the heart and pulmonary blood vessels than the femoral vein, there is an increased chance that blood drawn from this site may have been affected by postmortem diffusion from the heart.

The heart is often a source of a plentiful supply of

blood. However, if heart blood is collected and analyzed, peripheral blood samples must also be collected and analyzed to demonstrate whether significant differences in concentrations are present. The lungs are served with a plentiful supply of blood that can contain drugs in very high concentrations. After death, diffusion of drugs may occur from the lung into the left ventricle, resulting in blood drug concentrations in the left ventricle that are up to 10 times higher than those in the right ventricle. Particular caution should be exercised in the interpretation of results when blood samples have been collected from babies. In small infants, the volume of available blood may be limited because of the body size. There is also an increased risk that samples may contain blood drawn from the major blood vessels in the torso, where the risk of contamination by hepatic blood is high. In these cases consideration must be given to the fact that hepatic blood may contain drugs in concentrations that are orders of magnitude higher than in peripheral blood samples.

This reinforces the necessity for toxicologists to insist that all samples are clearly labeled with respect to the site and time of sampling. Where the site of sampling is in doubt, this should be stated in the toxicology report and the interpretation should be covered by appropriate caveats. Fluid scooped from the body cavity is occasionally submitted under the description of 'blood'. Such a sample is unsuitable for interpretative purposes, as it is seldom an authentic blood sample and there is a high probability that it is contaminated with stomach content, liver blood, bile, urine and intestinal contents.

If significant variations in drug concentrations may be encountered in blood drawn from different sites in the body, the blood drug concentrations published in the literature and used in reference collections may also have been subject to postmortem change phenomena. Data compilations of blood drug concentrations rarely provide detailed case histories and many were compiled before the extent of the implications of postmortem change were realized. Those tasked with interpreting postmortem blood concentrations should therefore exercise extreme caution before finally expressing their views in a report or before a court of law. Caution must also be exercised when, in the case of an extremely decomposed or an exhumed body, no blood is available and tissue drug concentrations are used for interpretative purposes. The practice of attempting to deduce a postmortem blood drug concentration from a tissue concentration by using tissue:blood ratios published in data collections is scientifically unsound and should not be undertaken.

Postmortem artifacts may also influence the ratios of metabolites to their parent drug in blood. For example, it has been suggested that the ratios of unconjugated to conjugated morphine might assist in indicating whether death had occurred within a short time of the entry of the drug into the body. This philosophy proposed that, if death occurred rapidly after entry of morphine or heroin into the body, the concentration of unconjugated morphine present in blood would be higher than the conjugated morphine concentration. The volume of distribution of morphine is, however, large compared with that of morphine glucuronides. Hence, while only a small proportion of morphine is present in the circulating blood, a significant proportion of the morphine glucuronides are present in blood. Any factors that could change the distribution of morphine, or materialize the conversion of some or all of the morphine glucuronides back into morphine, could grossly affect the interpretation of results based on morphine:morphine glucuronide ratios and subsequently provide an erroneous conclusion with respect to time of dosing in relation to death.

A list of concentrations for commonly encountered drugs or drug metabolites that have been widely associated with fatal poisoning is provided as **Table 1**. It must always be remembered, however, that in many cases blood drug concentrations cannot be interpreted in isolation and the task of interpretation should only be undertaken after full discussion has taken place between the toxicologist and the pathologist. Most experienced toxicologists will have encountered cases where so called 'fatal blood concentrations' have been measured in a deceased person but the cause of death was not directly associated with the drug detected. Drug addicts and palliative care patients can build up considerable tolerance to drugs such as opiates and appear to behave relatively normally, while the concentrations of opiates in their blood would be capable of causing death in a nontolerant subject. It is also well known that in many cases the cause of death may be directly attributed to gunshot wounds, stabbing or road traffic accidents but the concentrations of drugs in blood fall within the range of concentrations associated with fatal poisoning.

Our knowledge of postmortem change and the effects that may be associated with it is still far from being complete. However, an awareness that such phenomena exist should provide a warning to toxicologists and pathologists that the process of interpreting postmortem blood results is not simple. Because of the problems associated with drug tolerance and postmortem movement of drugs, all that can be safely deduced from a blood drug concentration alone is that the drug or its metabolites are present. A more elaborate interpretation should only be reached

Table 1 Postmortem blood concentrations of 50 common drugs and poisons that may be associated with fatal poisoning

Drug	*Concentration range associated with fatal poisoning* (μg ml^{-1})[a]
Acebutolol	35
Amitriptyline	2–16
Amphetamine	0.5–16
Amylobarbitone	20–30
Arsenic	0.6–9.3
Atropine	>0.2
Barbitone	>250
Butobarbitone	30–80
Caffeine	55–80
Carbon monoxide	$>45\%$ saturation
Chlormethiazole	15–60
Chloroform	10–48
Chlorpheniramine	>1.0
Chloroquine	1–125
Chlorpromazine	1–12
Cocaine (BE)	>0.9
Codeine	1–8.8
Cyanide	1.1–53
Cyclizine	>15
Desipramine	2.7–17
Dextropropoxyphene	1–60
Digitoxin	>0.004
Digoxin	0.005
Dihydrocodeine	0.8–17
Dothiepin	1.5–20
Doxepin	0.3–6
γ-Hydroxybutyrate (GHB)	27–221[b]
Glutethimide	40–75
Imipramine	0.3–30
Lignocaine	11
Lithium	>2.5
Methadone	>0.44
Methamphetamine	>40
Morphine	>0.5
Nortriptyline	0.3–26
Paracetamol	53–1500
Paraldehyde	>500
Pentazocine	>0.3
Phenobarbitone	>60
Phenytoin	>50
Promethazine	>2.5
Propranolol	2–4
Quinalbarbitone	5–52
Quinidine	30–50
Quinine	3–12
Salicylic acid	500
Thioridazine	>7
Trimipramine	>2.6
Warfarin	100
Zopiclone	0.9–3.9

BE, benzoylecgonine.

[a] The data listed should only be used after consideration of all of the caveats expressed in the text of this article.

[b] Postmortem blood concentrations of GHB may be unreliable as a result of postmortem production of GHB. It is recommended that urine should also be analyzed in addition to blood and, failing the availability of urine, vitreous humor. Blank postmortem blood samples have been reported to contain endogenously produced GHB at concentrations up to 107 μg ml^{-1}.

after all of the factors in a case have been drawn together, including the histological evidence, the gross pathology, the known circumstances of the case and the medical history of the deceased.

See also: **Drugs of Abuse**: Antemortem Blood.

Further Reading

Couper FJ and Logan BK (1999) Determination of GHB in clinical and postmortem samples. *Proceedings of the Society of Forensic Toxicologists Meeting, 1999*. (in press).

Mayer JM, Kofoed J and Robinson DW (1987) Postmortem changes in drug concentrations in the Blood. *Proceedings of the 24th International Association of Forensic Toxicologists Meeting*, Banff, Canada, pp. 40–47.

Moriya F and Hashimoto Y (1998) Redistribution of basic drugs into cardiac blood from surrounding tissues during early stages postmortem. *Journal of Forensic Sciences* **44**(1):10–16.

Osselton MD (1985) Toxicological tables – a compendium of pharmacological, therapeutic and toxicological data on 137 drugs and chemicals in man. In: Sunshine I (ed.) *Methodology for Analytical Toxicology*, vol. 3. Boca Raton, FL: CRC Press.

Pounder DJ and Davies JI (1994) Zopiclone poisoning: tissue distribution and potential for postmortem diffusion. *Forensic Science International* **65**:177–183.

Pounder DJ and Jones GR (1990) Post-mortem drug redistribution: a toxicological nightmare. *Forensic Science International* **45**:253–263.

Pounder DJ, Owen V and Quigley C (1994) Postmortem changes in blood amitriptyline concentration. *American Journal of Forensic Medicine and Pathology* **15**(**3**):224–230.

Pounder DJ, Fuke C, Cox DE, Smith D and Kuroda N (1996) Postmortem diffusion of drugs from gastric residue. *American Journal of Forensic Medicine and Pathology* **17**(1):1–7.

Prouty RW and Anderson WH (1990) The forensic science implications of site and temporal influences on postmortem blood drug concentrations. *Journal of Forensic Sciences* **35**(**2**):243–270.

Repetto MR and Repetto M (1997) Therapeutic, toxic and lethal concentrations in human fluids of 90 drugs affecting the cardiovascular and hematopoietic systems. *Clinical Toxicology* **35**(**4**):345–351.

Skopp G, Lutz R, Gan'mann B, Mattern R and Aderjan R (1996) Postmortem distribution pattern of morphine and morphine glucuronides in heroin overdose. *International Journal of Legal Medicine* **109**:118–124.

Spiehler V and Brown R (1987) Unconjugated morphine in blood by radioimmunoassay and gas chromatography/mass spectrometry. *Journal of Forensic Sciences* **32**(**4**):906–916.

Stephens BG, Coleman DE and Baselt RC (1999) In vitro stability of endogenous gamma-hydroxybutyrate in postmortem blood. *Journal of Forensic Sciences* **44**:231.

Urine

M-C Huang, B-L Chang and **C H Liao**, Department of Health, Taipei, Taiwan
R H Liu, University of Alabama, Birmingham, AL, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0491

Introduction

Analyses of drugs of abuse in biological media have been mainly associated with death investigation and criminal prosecution of those accused of driving under the influence of alcohol. In the United Sates of America (US), urine drug testing has recently become widely used as a deterrent tactic in combating drug abuse: in the military, the criminal justice system, and later in the workplace. In the aviation industry, urine drug testing is a common practice worldwide. As a result of the North American Free Trade Agreement, workplace urine drug testing is now extended beyond the US border. Recently, the government of Taiwan has also implemented urine drug testing programs in schools, the workplace and in the criminal justice system. These practices have dramatically increased the number of tests and the use of urine as the test specimen.

The *Mandatory Guidelines for Federal Workplace Drug Testing Program*, published by the US Department of Health and Human Services on 11 April 1988, and its updated versions, formulated the mechanisms by which the National Laboratory Certification Program oversees civilian drug testing activities in the US. These guidelines, and previously established military directives, adopt urine as the routine specimen for testing the following drug categories: amphetamines, barbiturates, benzodiazepines, cannabis, cocaine, lysergic acid diethylamide, methadone, methaqualone, opiates, phencyclidine and propoxyphene. (To the authors' knowledge, no random urine drug testing program included all of these drug categories.)

Test methodologies adopted by these testing programs normally include an immunoassay (preliminary or initial test), followed by confirmatory gas chromatography–mass spectrometry (GC-MS) analysis of the drug tested positive by the immunoassay. In a workplace drug testing program, test results are initially considered (within the laboratory) presumptive 'positive' or 'negative' – terms that do not hold the same meanings as commonly defined in the laboratory for scientific measurements. Universal administrative cutoff values, rather than scientific and statistical detection limit data established by individual laboratories, are used as the basis for determining whether a specimen is (presumptive) positive or negative (of the preliminary test). Those considered positive are further tested using a GC-MS protocol, again using a universal administrative cutoff concentration of the targeted analyte as the basis for deciding whether a specimen can be reported as positive. Cutoff values adopted by the US workplace and the military drug testing programs are shown in **Table 1**.

Urine Specimen Characteristics and Specimen Integrity

Specimen characteristics

Compared with blood, urine is relatively free of protein, thus making it possible for direct extraction. Urine also contains high concentrations of metabolites and provides a longer detection time for most drugs. Thus, urine is a preferred specimen for monitoring drug exposure (within a certain time window) of a large population on a regular basis.

Since certain drugs may not be found (or detected in low levels) in urine, while found in liver and blood samples in intoxicated concentrations, urine screen alone may not be sufficient for systematic toxicological analysis. Furthermore, blood is a better specimen for impairment or toxicological assessment purposes, as it is most closely related to drug concentrations at the effector sites.

Urine drug testing results depend on factors such as the time lapse between specimen collection and drug intake, and urine pH. For example, diet that causes urine pH changes may significantly alter the excretion patterns of certain drugs. Thus, weakly basic drugs (such as amphetamine) are more efficiently excreted in acid urine, whereas weakly acidic drugs (such as barbiturates and salicylates) are more efficiently excreted in alkaline urine.

Specimen integrity

Factors that may compromise specimen integrity, thus affecting test results, include:

- substitution of a nonurine liquid or urine from a different person;
- addition of an adulterant rendering the specimen unsuitable for certain test methodologies;
- the presence of certain characteristics associated with the specimen container or storage conditions.

Implementation of proper chain-of-custody procedures can best safeguard the integrity of a specimen. Laboratory procedures described later are often used to evaluate specimen integrity; they are, however, not always effective.

Table 1 Cutoffs[a] of immunoassay and GC/MS test adopted by the US certification programs

	Immunoassay (ng ml^{-1})			*GC-MS (ng ml^{-1})*		
Drug	*Analyte targeted*	*DoD*	*HHS*	*Analyte targeted*	*DoD*	*HHS*
Amphetamines	Amphetamines	–	1000			
	Amphetamine	500	–	Amphetamine	500	500
	Methamphetamine	500	–	Methamphetamine	500	500[a]
Barbiturates	Secobarbital	200	–	Amobarbital	200	–
				Butalbital	200	–
				Pentobarbital	200	–
				Secobarbital	200	–
Cocaine	Benzoylecgonine	150	300	Benzoylecgonine	100	150
Marijuana	9-THC-COOH	50	15	9-THC-COOH	50	15
Opiates	Morphine	300	300	Codeine	2000	2000
				Morphine	4000	2000[b]
				6-MAM	10	10[b]
Phencyclidine	Phencyclidine	25	25	Phencyclidine	25	25
LSD	LSD	0.5	–	LSD	0.2	–

DoD, US Department of Defense drug testing program; HHS, US Department of Health and Human Services, National Laboratory Certification Program; 9-THC-COOH, 11-nor-Δ^9-tetrahydrocannabinol-9-carboxylic acid; 6-MAM, 6-monoacetylmorphine; LSD, lysergic acid diethylamide.
[a] For administrative purposes, a sample is considered negative if the concentration of the target analyte is lower than the listed 'cutoff' value. It is also commonly practiced that a sample is considered positive only if $\geqslant$200 ng/ml amphetamine is also present.
[b] Testing of 6-MAM is required only when the concentration of morphine is $\geqslant$2000 ng ml^{-1}.
Courtesy of the American Association for Clinical Chemistry.

Specimen origin identification Analysis of blood group antigens and polymorphic proteins may help exclude a specific person as the specimen donor. These approaches are often ineffective owing to their low discrimination power and the large sample size required.

DNA-based technologies can be more effective. Squamous and transitional epithelial cells from the urinary tract often provide adequate DNA for specimen identification purposes. Since restriction fragment length polymorphism methods require larger DNA fragments and sample size, methods based on polymerase chain reaction offer higher success rates. Various genotypes of six markers that can be determined by a commercial kit are shown in **Table 2.**

Specimen adulteration and evaluation parameters Specimens can be adulterated *in vivo* through the use of diuretics (or simply a large volume of liquid), pH-altering liquids, or materials that cause the excretion of test-interfering substances. *in vitro* adulterations using household products are most widely reported events. These products often affect immunoassay test results. Some adulterants may actually destroy analytes, resulting in negative GC-MS test results.

It is often difficult to distinguish between an adulterated urine and a genuine one that responds marginally towards tests implemented to monitor parameters indicative of adulteration, such as specific gravity, pH and creatinine level. Selecting appropriate cutoff values of these parameters, that can best identify presumptive positive adulteration specimens for further confirmatory identification of adulterant, is no trivial matter. For example, adopting 1.007–1.035 specific gravity range and $>45\,mg\,dl^{-1}$ creatinine content as integrity criteria resulted in 69 (47.9%) out of the 144 highly suspicious urine specimens evading detection.

Table 2 Genetic markers and allele frequencies[a] using the PM+DQA1 typing kit

Marker	*Allele*	*Frequency*[a]
HLA DQA1	1.1	0.158
	1.2	0.190
	1.3	0.073
	2	0.145
	3	0.192
	4.1	0.214
	4.2/4.3	0.028
Low-density lipoprotein receptor (LDLR)	A	0.448
	B	0.552
Glycophorin A (GYPA)	A	0.530
	B	0.470
Hemoglobin G gammaglobin (HGBB)	A	0.537
	B	0.450
	C	0.013
D7S8	A	0.610
	B	0.390
Group specific component (GC)	A	0.275
	B	0.178
	C	0.547

[a] Data for US Caucasian males. Data taken from the reagent package insert (Perkin-Elmer, Foster City, CA).
Courtesy of Dr Alan HB Wu and *Forensic Science Review*.

Obviously, the most effective evaluation parameter varies with the adopted method of adulteration. Indicative parameters include temperature, color and smell, pH, and contents of additional substances, including diuretics, household items, and chemicals such as nitrite and glutaraldehyde reportedly being specifically formulated for urine adulteration purposes.

Analyte stability Stability of the analytes during the postcollection period may affect the quantitative data and the interpretation of these results. Important factors affecting the stability of an analyte include pH variations caused by intrinsic and extrinsic specimen characteristics; container properties and other storage conditions such as light and temperature; and the presence of oxygen, preservatives and other foreign substances.

Application of Immunoassays to Urine Drug Testing

Immunoassays are adopted as the 'official' preliminary test method in workplace urine drug testing programs because they are sensitive (without requiring pretreatment), can be automated and are based on an underlying principle that is different from the confirmatory GC-MS methods.

Common immunoassays

Most immunoassay procedures adopt the competitive binding principle, in which the antibody is allowed to react with a mixture of labeled (control) and unlabeled (sample) drugs. The presence and the quantity of a drug in the sample is evaluated on the basis of the quantities of the labeled drug in the reacted or unreacted forms. The control drugs are labeled in different ways, each requiring different methods of detection and quantification. For example, radioactive isotopes, such as ^{125}I, are used for labeling in radioimmunoassay (RIA) methods; active enzymes, which are capable of converting (indirectly) nicotinamide adenine dinucleotide (NAD) to its reduced form (NADH), are coupled to the drug in enzyme-multiplied immunoassay techniques (EMIT); fluoresceins are coupled to the drug used in fluorescence polarization immunoassay (FPIA); and particles are attached to the drug in particle immunoassay (PIA).

An immunoassay system is considered heterogeneous if a phase-separation step is needed prior to detection. The detection process is designed to measure the extent of the labeled antigen linked (directly or indirectly) to the antibody, thus reflecting the amount of the test drug (unlabeled antigen) in the sample. Methodologies based on the measurement of radioactivity are heterogeneous because the detecting device cannot differentiate the source of the radioactivity (free or bound labeled antigen). Therefore, a separation step is required.

Methodologies based on a change in optical intensities do not require the separation step if these properties are modified through the substrate's linkage (directly or indirectly) to the antibody. They are considered homogeneous immunoassays. Both heterogeneous and homogeneous immunoassays are used for workplace drug urinalysis.

Interferences and crossreactivity

Interference can be broadly defined as the cause for a test result that does not provide the intended diagnostic finding reflecting the true status of the specimen. The following conditions may generate test results leading to incorrect interpretation of the specimen status:

- the presence of the targeted analyte derived from sources other than the targeted drugs of abuse;
- the presence of crossreacting compounds with known or unknown structures;
- specimen conditions that cause nonspecific binding or interfere with the assay's detection mechanism;
- the presence of adulterants that degrade the analytes or alter their interacting characteristics.

Targeted analytes from legal sources Some of the analytes targeted as indicators of drug abuse may derive from unintended exposure, food consumption or licit medication. For example, it is well known that morphine and codeine may be observed in urine specimens collected from individuals consuming food items with poppy seeds or prescriptions containing morphine or codeine. Methamphetamine and amphetamine can derive from the use of Vicks nasal inhaler and a substantial number of licit drugs.

Crossreacting compounds Crossreacting compounds are widely reported and many are listed in reagent inserts or documentation provided by the manufacturers. Identities of most reported crossreacting compounds are known, while others are not. For example, it has been reported that unknown metabolites of chlorpromazine, brompheniramine and labetalol caused EMIT d.a.u. Monoclonal Amphetamine/Methamphetamine Assay to generate false-positive results. Metabolites, not the parent drugs, are believed to be the responsible crossreacting compounds because:

- these drugs were prescribed for the urine specimen donors;

- these parent drugs were present in the urine specimens;
- studies on control samples with various concentrations of the parent drugs *alone* failed to generate a positive result.

Since reference metabolites of these drugs are not available, the exact crossreacting metabolites cannot be identified.

Nonspecific binding Many postmortem urine specimens reportedly produced absorbance change values lower than those resulting from negative calibrators and specimens collected from healthy persons. This is attributed to the presence of nonspecific interacting materials that cause a higher initial absorbance value.

Detection mechanism Enzyme immunoassay methods are most likely to suffer spectrometric interference caused by other substances present. Listed below are some examples:

- *p*-Nitrophenol (a metabolite of methyl parathion) and tolmetin (a nonsteroidal anti-inflammatory drug) and its metabolites can absorb strongly in the 340 nm region at pH 8.0 and thus cause interference.
- The presence of metronidazole or mefenamic acid causes excessively high initial absorbance values, thus preventing the assessment of EMIT test data.
- Salicyluric acid (the principal metabolite of aspirin) interferes with the EMIT test methodology by reducing the molar absorptivity of NADH at 340 nm.

Adulterants Adulterants may actually destroy the targeted drugs, thus rendering the specimen 'truly' negative, while most adulterants cause nonspecific binding or create interference to the targeted antibody–antigen reaction or the detection mechanism.

Many household products are known to affect the responses of common immunoassays. Results from an exemplar study are shown in **Table 3**. Most interference studies did not compare the effects of adulterants on various immunoassays under the same conditions; it is therefore difficult to make general statements concerning the robustness of one methodology over the others. It seems to be clear, however, that cannabinoid assays are more susceptible to the interfering effects of adulterants.

Numerous mechanisms have been proposed to account for the observed interference, but exact causes are generally unknown. It has been proven, however, that bleach actually causes the degradation of 9-THC-COOH. Visine was believed to increase the adhesion of 9-THC-COOH to the borosilicate glass specimen containers, thereby reducing the availability of 9-THC-COOH in antibody-based assays.

Application of GC-MS in Urine Drug Testing

Confirmatory test procedures for drugs of abuse in biological media often include acid, base or enzyme hydrolysis; liquid–liquid or solid-phase extraction; and often with derivatization. Selective ion monitoring (SIM) GC-MS protocols are used almost universally for the analysis of the limited number of drugs/metabolites targeted by workplace drug testing programs.

Sample pretreatment

Many free drugs (and their metabolites) have strong affinity to proteins, whereas most conjugates are highly hydrophilic and not amenable to extraction by organic solvents. Thus, pretreatment procedures generally include the hydrolysis of drug conjugates (if present), the removal of the binding proteins, and the extraction of the drugs from the resulting reaction mixture. Since urine has limited protein content, protein removal is generally not a critical step.

Acid and enzyme hydrolysis of conjugated drugs Both acid hydrolysis and enzyme digestion have their merits. For example:

- β-glucuronidase is known to release intact benzodiazepines from their glucuronide conjugates without converting to their corresponding benzophenones, which is a typical product of acid hydrolysis.
- compared to β-glucuronidase, acid hydrolysis of morphine glucuronide is less time-consuming and can achieve a higher recovery rate. However, this process tends to cause the decomposition of acid-labile compounds in the matrix, leaving a dirty reaction mixture.

Enzyme hydrolysis may be preferred when the analyte is acid-labile and high recovery is not an important factor, i.e. when sensitive detection and quantification methodologies are incorporated in the overall protocols.

Extraction An extraction step typically removes interfering materials and concentrates the analyte (through reconstituting the extraction residue into a small volume of suitable solvent prior to final instrumental analysis).

Liquid–liquid extraction systems are designed to have the analyte partition preferably into the organic

Table 3 Effect of various adulterants on immunoassays for drugs of abuse

Adulterant	*Amphetamines*	*Barbiturates*	*Benzodiazepines*	*Cocaine*	*Opiates*	*Phencyclidine*	*Marijuana*
Ammonia							+R
Ascorbate							−R
Bicarbonate	−R	−R			−E	−F	
Bleach	−R; −E; −F; −C	−C	−E; −C	−C	−R; −E; −F; −C	−R; −E; −F; −C	−E; −F; −C
Detergent	−C	+F; −C	−C	+F; −C	−C	−C	
Drano	−E; −C	−E; −C	−E; −C	−R; −E; −F; −C	−E; −C	−C	−R; −E; −C
Golden seal	−C		−F				−R; −E; −C
Hand soap	+F	+F	−E				−R; −E
Joy	+F	+F	+R; −E;+F		−E		+R; −E;+F
Lime	−R				−R		−R
Peroxide			−E;+F				
Phosphate				−R−F	−F	−R	
Salt (NaCl)	−E	−E	−E	−E	−E	−E	−E
Vanish	−R				−R		−R
Vinegar							−E; −F
Visine			−E				−R; −E; −C

Adulterant reduces (−) or increases (+) response to drug listed by methodology. R, radioimmunoassay; E, enzyme multiplied immunoassay technique; F, fluorescence polarization immunoassay; C, cloned enzyme donor immunoassay.
Courtesy of Dr Alan HB Wu and *Forensic Science Review*.

phase in its unionized form or as an ion pair. The analyte may also be 'salted out' of the aqueous phase into the organic one by greatly increasing the ionic strength of the former using a high concentration of electrolytes. pH adjustment of the sample or the addition of a proper counterion may be used to facilitate a favorable partitioning of the drug in its unionized form or as an ion pair, respectively.

In a solid-phase extraction process, separation is performed with a solid additive or column material using hydrophilic, hydrophobic or ionic groups attached to silica. Solid-phase extraction approaches offer the following advantages over conventional liquid–liquid procedures:

- less organic solvent usage;
- no foaming problems;
- shorter sample preparation time;
- ease of incorporation into an automatic operation process.

Solid-phase extraction procedures now widely adopted for the analysis of drugs of abuse in urine include four basic operation steps: conditioning, loading, rinsing and eluting. The first solvent for conditioning should be as strong as, or stronger than, the elution solvent. The second conditioning solvent should be the same as, or as close to the strength of, the loading solvent as possible. The solvent used for loading should be as weak as possible to produce the tightest or narrowest band of adsorbed sample on the sorbent. A solvent that is slightly stronger or the same strength as the loading solvent is used as the rinse solvent. The rinse will elute unwanted sample components that are not as strongly retained as the analytes, and also wash down small droplets of loading solvent adhering to the walls of the tube to ensure that all sample comes in contact with the sorbent. An ideal eluting solvent should elute the analytes within 5–10 bed volumes. The optimal amount of solvent to elute the analytes from a 500 mg cartridge is about 0.6–1.2 ml. Using a solvent that is too strong will result in the elution of unnecessary sample components that are more strongly retained than the analytes, whereas a solvent that is too weak will result in excessive elution solvent volumes, which negate the advantage of reducing solvent consumption with solid-phase extraction cartridges. Sometimes, a desired solvent strength may have to be obtained by blending appropriate amounts of miscible solvents.

If a water-immiscible eluant is selected and the final analysis is to be performed with GC, a 'drying' procedure should be applied between the rinsing and the eluting steps. It has been reported that the combination of vacuum and a small amount of methanol can produce a 'dry' eluate without causing a substantial loss of drugs.

Chemical derivatization

Many drugs/metabolites are derivatized, prior to the measurement step, to bring the analytes to the chemical forms that are compatible with the chromatographic environment or to maximize their chromatographic separation and detection efficiencies. For example, derivatizations are carried out to improve the analyte's volatility; to eliminate active functional groups that may cause undesired interactions with the chromatographic components, resulting in peak loss (due to irreversible adsorption) or peak tailing (caused by reversible adsorption); and, through the use of a chiral agent, to achieve resolution of enantiomeric components.

Derivatization is also an effective approach to achieve higher detection or quantification efficiency or to facilitate analyte structure elucidation. For example, the hydroxy group in oxazepam is derivatized to prevent ring contraction that may occur at elevated temperatures. In GC-MS applications, mass shifts in the spectra produced by different derivatizing agents can reveal valuable information concerning the functional groups in the analyte.

Qualitative analysis

Serving as a chromatograph detector, a master spectrometer (MS) provides valuable 'fingerprint' information by resolving and displaying charged particles originated from chromatograph eluates. Together with the chromatographic retention data, the MS detection provides test results that are normally considered conclusive for analyte identifications. MS detector also facilitates adoption of an isotopic analog of the analyte as the internal standard in an analytical protocol.

Serving as an internal standard, an isotopic analog of the analyte incorporated at the very early stage of the analytical process serves as a model compound which, through its detection at the final step, provides a mechanism to prove successful completion of the entire analytical protocol. (Verifying a successful protocol is essential when the targeted analyte is not detected in the test sample.) Internal standards also provide intrasample retention time and ion intensity data that help analyte identification and quantification.

In general, fully credible qualitative analysis (as opposed to target-compound analysis) requires operating the MS in full-scan mode. The resulting spectra are then compared with those in the database or with

an in-house standard; together with the GC retention data, analyte identification is often conclusive.

Data collection using SIM mode can provide enhanced sensitivity and is commonly used as an integral part of a well-designed target compound analysis (qualitative and quantitative) scheme. Without the benefits of complete spectra, the use of SIM data for qualitative determination purposes should be exercised with extreme caution. For example:

- As many ions that are characteristic of the analyte should be selected. Ideally, the molecular ion should be used if it exists with a reasonable intensity. If the analyte is derivatized with a derivatizing reagent prior to the GC-MS analysis, at least one of the ions selected should include the complete or a characteristic moiety of the analyte. The exclusive use only of ions that are derived from the derivatizing reagent is not acceptable.
- The intensity ratios of the selected ions should be closely monitored and compared with the corresponding ratios obtained from a standard (or control) that is analyzed under identical conditions.
- The retention times of all ions monitored for the analyte should be coincidental. In addition, the retention time and ion intensity ratios of the internal standard should be compared to the corresponding data of the analyte. This 'intrasample' information further improves the certainty of an identification.
- This approach should only be applied to well-studied systems in which isotopic analogs of the analytes are normally available and incorporated in the analytical process. Furthermore, GC operation conditions should be optimized so that closely related compounds are chromatographed with distinguishable retention data.

Considering the fact that the identified analyte has also survived the chemical processes (hydrolysis, extraction and derivatization) designed for the targeted drug and proven effective for the internal standard (often a deuterated analog with chemical properties practically identical to the analyte), the above identification criteria should not reach an incorrect conclusion if executed by an able analyst.

Quantitative determination

Although MS by itself is not necessarily the most quantitative detector for a chromatographic system, the use of an isotopic analog as the internal standard alleviates the effects of many variations in the analytical process, thus often providing the best overall quantitative result.

Internal standard Deuterated analogs of the analytes are now commonly used internal standards for the quantification of drugs in biological matrices. However, not all deuterium-labeled isotopic analogs (of the analyte) are effective. Since the analyte and the internal standard are rarely separated adequately, the proposed isotopic analog must generate at least one (preferably two or three) ions relatively free of cross-contribution by the analyte. There must also be at least three ions designated for the analyte that are relatively free of cross contribution by the proposed internal standard. If these requirements are not met, quantitative results and the ion intensity ratio data (which are commonly used as important parameters for analyte identification) may become unreliable.

Calibration approaches A typical quantitative GC-MS protocol involves monitoring several selected ions from the analyte and the isotopic analog. Quantification is achieved by comparing a selected analyte-to-isotopic analog ion intensity ratio observed from the test specimen and the same ratio observed from the calibration standard. The calibration standard contains the same amount of the internal standard and a known amount of the analyte and is processed in parallel with the test specimen. The analyte concentration in the test specimen can be calculated using a one-point calibration approach, as shown in **Fig. 1**.

This procedure can be used to determine directly the analyte concentration in a test specimen (one-point calibration) and has proven highly effective when the concentration of the analyte is in the immediate vicinity of the single calibrator's concentration. Alternatively, the ion intensity ratio data can be used to determine the responses of a set of standards, from which a calibration curve is established

$$\frac{\left[\dfrac{(\text{Selected ion intensity})_{\text{Analyte}}}{(\text{Selected ion intensity})_{\text{Internal Std}}}\right]_{\text{Test sample}}}{\left[\dfrac{(\text{Selected ion intensity})_{\text{Analyte}}}{(\text{Selected ion intensity})_{\text{Internal Std}}}\right]_{\text{Calibration standard}}} \times \left[\begin{matrix}\text{Standard}\\ \text{conc.}\end{matrix}\right]_{\text{Calibration standard}} = \left[\begin{matrix}\text{Analyte}\\ \text{conc.}\end{matrix}\right]_{\text{Test sample}}$$

Figure 1 Formula for the calculation of the analyte concentration using a one-point calibration protocol. (Courtesy of Central Police University Press.)

and used for analyte concentration determination in a test specimen (multiple-point calibration). Multiple-point calibration approaches can be implemented with different regression models, including linear or nonlinear; weighted or unweighted; and whether or not to force the regression through the point of origin.

It has been shown that multiple-point nonlinear calibration can provide a wider calibration range. Empirical studies also indicated that one-point calibration quantification results for pentobarbital were actually inferior to those obtained for the other three barbiturates – even though d_5-pentobarbital was used as the single internal standard for all analytes. Factors that should be carefully evaluated include:

- cross contribution of quantification ions selected for the analyte and the internal standard;
- selection of an appropriate calibration model to take into account the cross contribution interference – a common phenomenon.

Instrumental parameters For best accuracy and precision of SIM measurements, instrumental parameters, such as instrument resolution, ion-monitoring position, threshold setting and dwell time, must also be carefully considered. Instrument drift from the peak center and the fluctuations in intensities of the selected ion beams will affect measurement precision, mainly because of the greater variation in the intensities of the weaker signals detected. Improper threshold settings will also affect the weaker signals to a greater extent. To achieve the best results, the signal level of the internal standard should be comparable to that of the analyte.

For GC-MS applications, problems associated with differences in ion residence time in the ion source must also be addressed. To assure adequate accuracy, the number of data points that define an ion chromatographic peak must be sufficient, preferably about 20. Because ions are sequentially monitored, the number of data points for a GC peak depends not only on the chromatographic peak width, but also on the ion-monitoring cycle time. The cycle time depends on the number of ions monitored, the time spent on monitoring each ion (dwell time) and the 'overhead time' needed by the system for effective switches.

Application examples

The implementation of workplace testing programs prompted the development of many robust extraction–derivatization GC-MS procedures for routine and large-scale analysis of targeted drugs in urine specimens. Examples of these protocols are summarized in **Fig. 2**. Evaporated residues are reconstituted and analyzed using GC-MS parameters shown in **Table 4**. Qualitative and quantitative determination criteria described in earlier sections are used for these procedures.

Correlation of Immunoassay and GC-MS Test Results

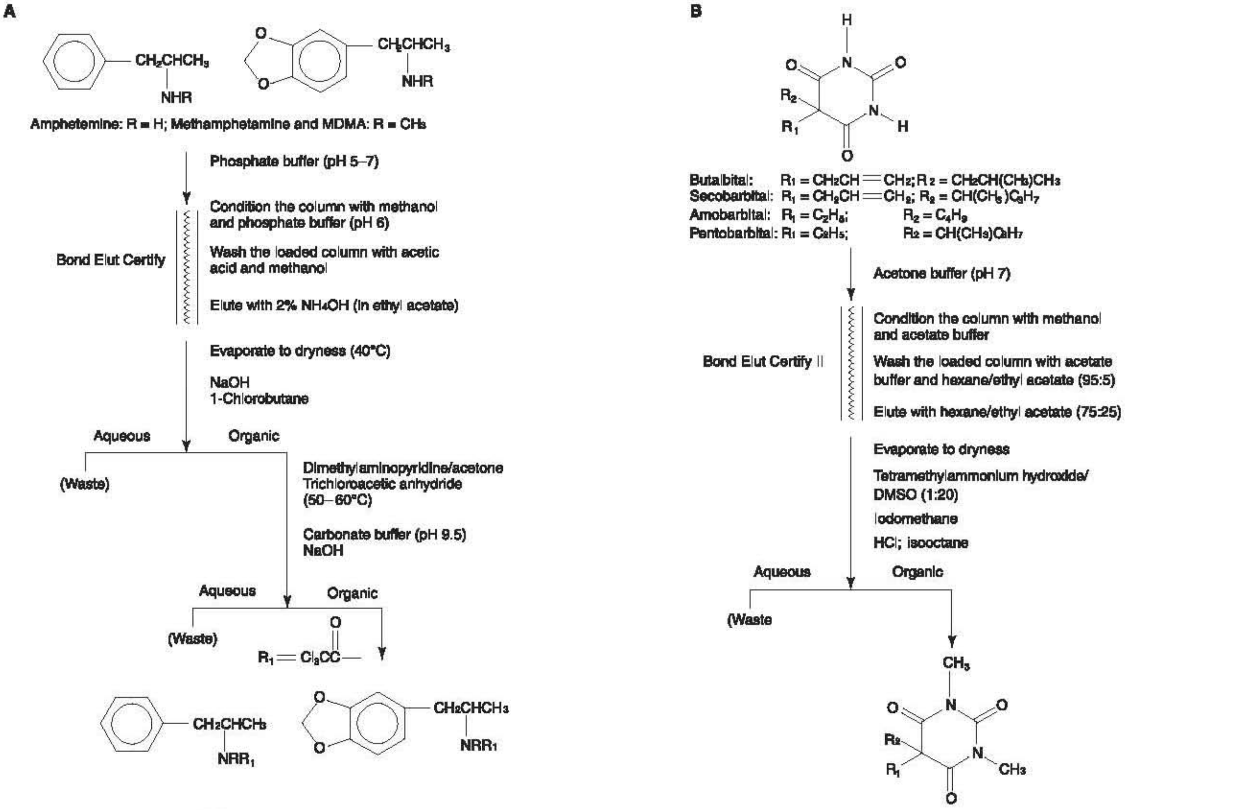
Conventional immunoassay reagents are typically not very specific. In fact, crossreactivity to drugs with similar structures can be advantageous in general applications. It is, however, a significant disadvantage when used in workplace drug testing programs, in which a positive result is reported only if the GC-MS result (of a targeted drug/metabolite) is at or above a cutoff concentration, and the prior immunoassay screen test result is also at or above an identical or higher cutoff level. Thus, manufacturers often strive to reduce a reagent's crossreacting characteristics and improve the detection mechanism so that immunoassay test results can be more comparable with GC-MS, thereby improving the overall test efficiency. The ideal situation is that all samples screened positive are confirmed to contain the target drug/metabolite at or above the GC/MS cutoff level, while those screened negative contain no (or below GC/MS cutoff) targeted drug/metabolite.

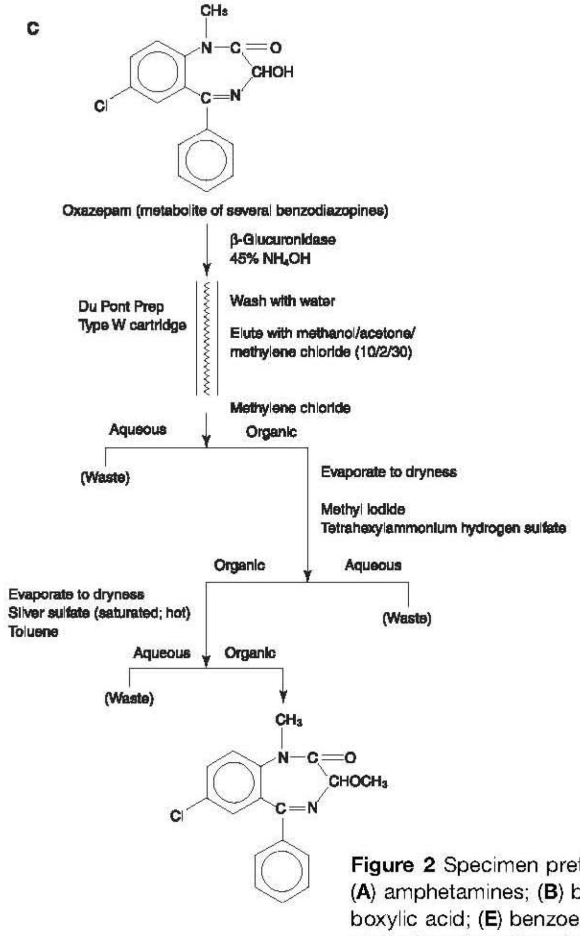
Because immunoassay reagents are not absolutely specific, the ideal situation can never be realized. One can hope, however, for an immunoassay to respond sensitively toward variations of the analyte content, especially in the vicinity of the selected cutoff concentration. Furthermore, the immunoassay responses should correlate reasonably with the GC-MS results. The appropriate relationship between the cutoffs set for these two tests depends on the metabolites distribution pattern of the drug and the crossreactivity characteristics of the adopted immunoassay.

Testing for marijuana use serves as an excellent example to illustrate the importance of cutoff selection when using an immunoassay as the preliminary test for GC-MS analysis. With a specific cutoff adopted for the GC-MS test, overall positive rate varies with the cutoff level adopted for the immunoassay. Immunoassay cutoffs that best correspond to a specific GC-MS cutoff varies with the type, the manufacturer and the lot of the adopted immunoassay.

Issues in the Interpretation of Urine Drug-Testing Results

It has been widely publicized that urine specimens tested positive for opiate and amphetamine may actually be caused by the use of licit medication or food items. For the opiate drug category, it is possible





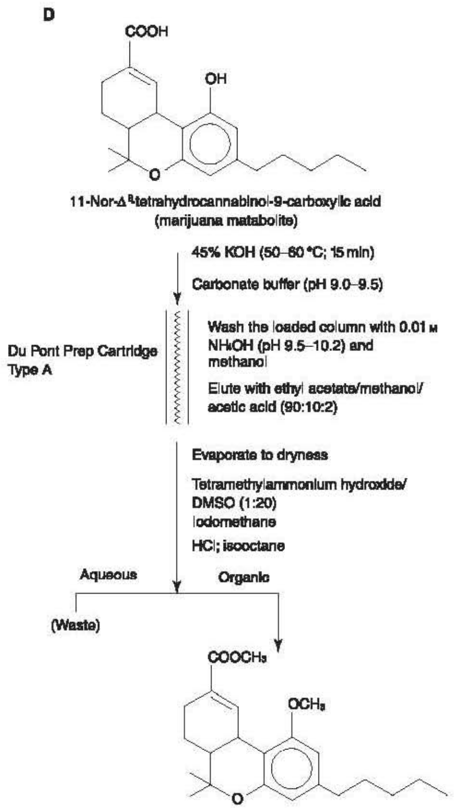
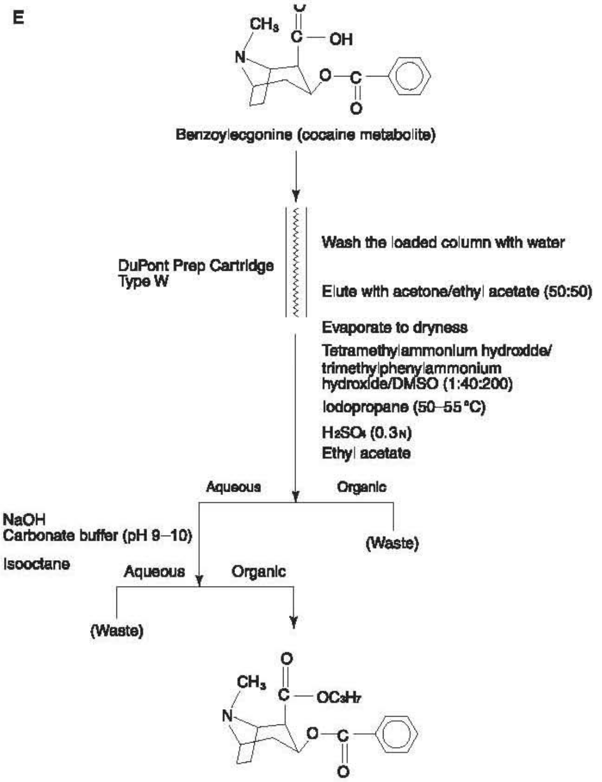


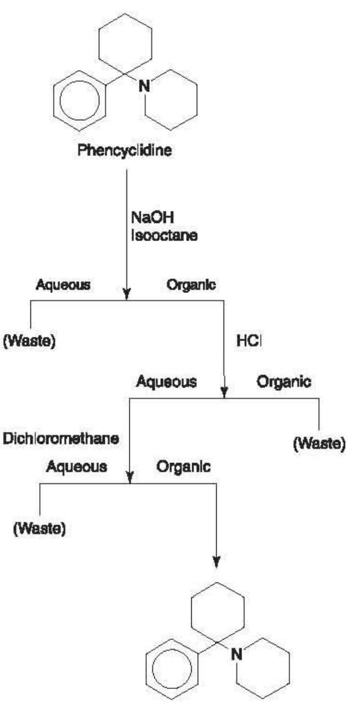
Figure 2 Specimen pretreatment procedures for GC-MS analysis of selected drugs in urine: (**A**) amphetamines; (**B**) barbiturates; (**C**) oxazepam; (**D**) 11-nor-Δ^9-tetrahydrocannabinol-9-carboxylic acid; (**E**) benzoecgonine; (**F**) phencyclidine; (**G**) methadone; and (**H**) opiates. (Courtesy of the American Chemical Society.)

E



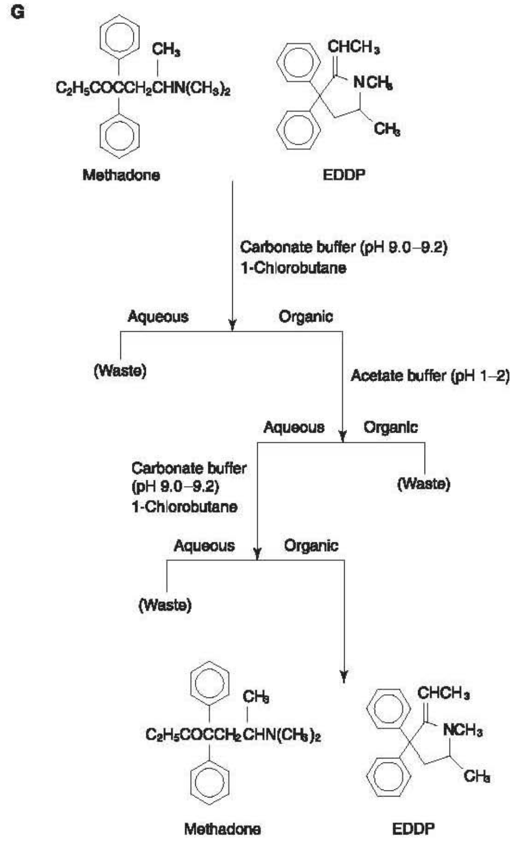
Benzoylecgonine (cocaine metabolite)
DuPont Prep Cartridge Type W
Wash the loaded column with water
Elute with acetone/ethyl acetate (50:50)
Evaporate to dryness
Tetramethylammonium hydroxide/ trimethylphenylammonium hydroxide/DMSO (1:40:200)
Iodopropane (50–55 °C)
H_2SO_4 (0.3 N)
Ethyl acetate
Aqueous
Organic
NaOH
Carbonate buffer (pH 9–10)
Isooctane
(Waste)
Aqueous
Organic
(Waste)

F



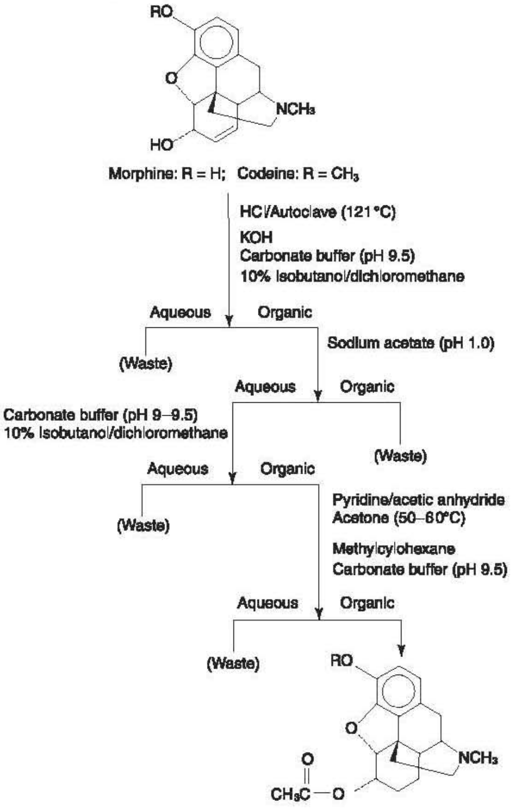
Phencyclidine
NaOH
Isooctane
Aqueous
Organic
(Waste)
HCl
Aqueous
Organic
Dichloromethane
(Waste)
Aqueous
Organic
(Waste)

G



Methadone
EDDP
Carbonate buffer (pH 9.0–9.2)
1-Chlorobutane
Aqueous
Organic
(Waste)
Acetate buffer (pH 1–2)
Aqueous
Organic
Carbonate buffer (pH 9.0–9.2)
1-Chlorobutane
(Waste)
Aqueous
Organic
(Waste)
Methadone
EDDP

H



Morphine: R = H; Codeine: R = CH_3
HCl/Autoclave (121 °C)
KOH
Carbonate buffer (pH 9.5)
10% Isobutanol/dichloromethane
Aqueous
Organic
(Waste)
Sodium acetate (pH 1.0)
Aqueous
Organic
Carbonate buffer (pH 9–9.5)
10% Isobutanol/dichloromethane
(Waste)
Aqueous
Organic
(Waste)
Pyridine/acetic anhydride
Acetone (50–60°C)
Methylcyclohexane
Carbonate buffer (pH 9.5)
Aqueous
Organic
(Waste)

Table 4 Important parameters for SIM GC-MS analysis of targeted drugs

	Targeted GC				
Targeted drug category and drug[a]	*Recon. solvent (approx. vol.)*	*Oven temp. range (°C)*	*Ions monitored (m/e)*[b] *Analyte; Int. Std*[c]	*Ion ratios monitored*	*Cutoff*[d] *(ng ml^{-1})*
Amphetamines	Ethyl acetate				
Amphetamine	(40–100 μl)	150–200	118, 188, *190*; 123, *194*	118/190, 188/190; 123/194	500
Methamphetamine			91, 202, *204*; 209, *211*	91/204, 202/204; 211/209	500
MDMA			204, *162*, 202; *164*, 208	204/162, 202/162; 208/164	500
l-Amphetamine	Ethyl acetate	100–250	*237*; *241*		500
d-Amphetamine	(10 μl)	100–250	*237*; *241*		
l-Methamphetamine		100–250	*251*; *258*		
d-Methamphetamine		100–250	*251*; *258*		
Barbiturates	Ethyl acetate				
Butalbital	(20 μl)	140–190	181, 195, *196*; 171, *189*	181/196, 195/196; 171/189	200
Amobarbital			*169*, 184, 185; 171, *189*	184/169, 185/169; 171/189	200
Pentobarbital			169, *184*, 185; 171, *189*	184/169, 185/169; 171/189	200
Secobarbital			181, 195, *196*; 171, *189*	153/168, 195/168; 171/189	200
Cannabis	Cyclohexane	220–270	313, 357, *372*; 360, *375*	313/372, 357/372; 360/375	15
9-THC-COOH	(20 μl)				
Cocaine	Cyclohexane	190–240	*210*, 226, 331; *213*, 334	210/331, 226/331; 213/334	150
Benzoylecgonine	(20 μl)				
Opiates	Ethyl acetate	200–270	229, 282, *341*; 232, *344*	229/341, 282/341; 232/344	2000
Codeine	(20 μl)		310, 327, *369*; 330, *372*	310/369, 327/369; 372/330	2000
Morphine					
Phencyclidine	Methanol	170–210	186, *200*, 243; *205*, 248	186/243, 200/243; 205/248	25
	(20 μl)				
Oxazepam	Cyclohexane		255, 271, *314*; 276, *319*	255/271, 314/271; 276/319	300
	(20 μl)				
Methadone	Ethyl acetate		*223*, 294, 295; *226*, 297	223/294, 295/294; 226/297	300
	(20 μl)				

MDMA, 3,4-methylenedioxymethamphetamine.

[a] Some of these drugs/metabolites were analyzed as derivatives using methodologies shown in Fig. 1. *N*-Trifluoroacetyl-*l*-prolyl chloride was used for enantiomeric composition analyses of amphetamines.

[b] Italic ions are used for quantification.

[c] 1-Phenyl-2-aminopropane-1,2,3,3,3-d_5, 1-phenyl-2-(methyl-d_3-amino)propane-1,1,2,3,3,3-d_6, 1-(3,4-methylene-dioxy)phenyl-2-(methylamino)propane-1,2,3,3,3-d_5, pentobarbital-5-ethyl-d_5, 11-nor-Δ^9-tetrahydrocannabinol-9-carboxylic acid-5′-d_3, benzoylecgonine-*N*-methyl-d_3, codeine-*N*-methyl-d_3, morphine-*N*-methyl-d_3, phencyclidine-phenyl-d_5, oxazepam-phenyl-d_5, and methadone-1,1,1-d_3 are used as the internal standards for respective analytes.

[d] For administrative purpose, a sample is considered negative if the concentration of the target analyte is lower than the listed 'cutoff' value. It is also common practice for a sample to be considered positive only if $\geqslant$200 ng ml^{-1} amphetamine is also present.

Courtesy of the American Chemical Society.

to confirm heroin use if 6-monoacetylmorphine is detected. Concentrations of morphine and codeine and their ratio can also provide useful, but often inconclusive, information.

Detection of methamphetamine as an analytical artifact has been attributed to the presence of high level of ephedrine or pseudoephedrine in urine specimens. It is thus recommended that a urine specimen cannot be reported positive for methamphetamine without the presence of amphetamine.

Amphetamine and methamphetamine may derive from the use of Vicks nasal inhaler and other medicines. For example, amphetamine is available as Dexedrine, which consists solely of dextroamphetamine (*d*-amphetamine), and benzphetamine, which consists of equal amounts of *d*-amphetamine and racemic (*d*,*l*-) amphetamine. Amphetamine and methamphetamine may also be present as metabolites of other drugs (precursors), such as amphetaminil, benzphetamine, clobenzorex, deprenyl, dimethylamphetamine, ethylamphetamine, famprofazone, fencamine, fenethylline, fenproporex, furfenorex, mefenorex, mesocarb and prenylamine. Since amphetamine and methamphetamine derived from these drugs have certain enantiomeric composition characteristics, enantiomeric analysis can often provide valuable information, but not always with definite conclusions.

Unintended passive inhalation of marijuana and cocaine can also result in the detection of the metabolites normally monitored to indicate the abuse of these drugs. The detection of cocaine metabolite has also been attributed to drinking Health Inca Tea, handling of cocaine-containing currency, and skin absorption for those who are exposed to cocaine at their work environment. Analyte concentrations resulting from drug exposure are generally low; claims should be evaluated carefully on a case-by-case basis. Opinions are often offered; definite conclusions are not always achievable based on test results alone.

Acknowledgments

We are grateful to the Taiwanese National Science Council for providing a visiting appointment (NSC87-2811-B-043A-0001) to Ray H Liu, which greatly facilitated the preparation of this publication.

See also: **Analytical Techniques**: Mass Spectrometry. **Drugs of Abuse**: Classification, including Commercial Drugs.

Further Reading

Department of Health and Human Services, Substance Abuse and Mental Health Services Administration (1994) Mandatory guidelines for federal workplace drug testing programs: notice. *Federal Register* **59**:29908–29931.

Edwards C, Fyfe MJ, Liu RH and Walia AS (1993) Evaluation of common urine specimen adulteration indicators. *Journal of Analytical Toxicology* **17**:251–252.

Liu RH (1994) *Elements and Practice in Forensic Drug Urinalysis*. Taipei: Central Police University Press.

Liu RH and Gadzala D (1997) *Handbook of Drug Analysis: Application in Forensic and Clinical Laboratories*. Washington, DC: American Chemical Society.

Liu RH and Goldberger BA (1995) *Handbooks of Workplace Drug Testing*. Washington, DC: American Association for Clinical Chemistry.

Liu RH, McKeehan AM, Edwards C *et al.* (1994) Improved gas chromatography/mass spectrometry analysis of barbiturates in urine using centrifuge-based solid-phase extraction, methylation, with d_5-pentobarbital as internal standard. *Journal of Forensic Science* **39**:1501–1514.

Wu AHB (1998) Integrity of urine specimens for toxicological analysis: adulteration, mechanism of action, and laboratory detection. *Forensic Science Review* **10**:47–65.

DUST

S Palenik, Microtrace, Elgin, IL, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0496

Introduction

Dust is the microscopic, particulate residue of a universe that is crumbling under the forces of nature and human beings and has been described by the great chemist Liebig as matter in the wrong place. Because microscopic particles are characteristic of their source, easily transferred between objects and persons, and as transfer is difficult to prevent, dust has special significance in forensic science. It was through study of and experimentation on dusts of all types that Dr Edmond Locard of Lyon laid the foundation for the collection, analysis and interpretation of contact traces in the first half of the twentieth century. The results of this work were formulated into what L C Nickolls called Locard's exchange principle: 'Whenever two objects come into contact there is always a transference of material.' This principle is the basis and justification for the detection, recovery and analysis of microscopic trace evidence. Since the quantity of matter which may have been transferred is frequently quite small, and since all transferred material is subject to loss from the surface to which it was

transferred, over time and through motion, these tasks must be performed by a trace evidence analyst who is equipped both by training and experience to perform these tasks in a thorough, scientific and efficient manner.

Goals and Uses of Dust Analysis

As one might expect, both the goals and techniques of dust analysis have changed since Locard's time. Locard's experiments were directed primarily toward identifying the occupation and geographic movements of a suspect from an analysis of the dust from clothing, earwax, fingernails, etc., and using this information to help solve crimes. In one case, for example, he isolated the dust from the clothing of suspected counterfeiters. Microchemical analysis showed that the metallic particles comprising the majority of the dust consisted of tin, antimony and lead, the very metals of which the counterfeit coins were composed. The suspects' normal occupations would not have produced particles such as those that were found on their clothing. Today's trace evidence analyst uses microscopic evidence to establish links in various combinations between a victim, suspect and crime scene. For example, fibers similar to those from the clothing of a suspect may be found on the body of a victim, or soil similar to that at the scene of the crime may be recovered from a suspect's shoes. Less often, the laboratory is asked to provide investigative leads to detectives in the field. In these cases, the scientist may be asked, for example, to determine the likely origin of particles found on the bodies of the victims of a serial killer. This was the case with colored paint spheres recovered from the naked bodies of murdered prostitutes in London between 1964 and 1965. These particles led the police to a repaint shop to which, as it turned out, the killer had access. In another example, the origin of sand, substituted for a rail shipment of silver flake between Colorado and New York, was determined, on the basis of a mineralogical examination of the sand grains, to have originated from a location near Denver.

One can see from the examples above that the analysis of microscopic particles has the potential to provide a wealth of useful information during the investigation of a crime and the trial that follows after a suspect has been identified and charged. The principal questions that can be addressed by the analysis of dust fall naturally into three main categories, although others may suggest themselves in certain circumstances.

1. *Is there evidence of likely contact, in some combination, between a victim, suspect and the scene of the crime*? This is the most commonly asked question and accounts for most courtroom testimony involving microscopic trace evidence. If Locard's exchange principle holds true, there must be a transfer of material if contact was made. It is important to remember, however, that the inability to detect physical evidence of a contact does not mean that no contact occurred. As someone once wisely said, 'Absence of evidence does not mean evidence of absence.' Even through a transfer of material occurs, the quantity may be so small as to escape detection, or the decay rate may be such that evidence is lost before it is brought to the laboratory for examination. Another possibility related to quantity is mixing. This is a particular problem with certain types of trace evidence, such as soil taken from the wheel well of a vehicle, where mixing of various soils may make it impossible to reach a conclusion regarding the identity of a source. Fibers are the type of particulate evidence most commonly collected and studied. This is due in great part to three factors. The first is the common occurrence of fibers in our modern environment from clothing, carpets and upholstery. The second is the publicity surrounding a number of high profile murder investigations that were solved and prosecuted primarily with the aid of fiber transfers, such as those associated with the Atlanta child murders in the United States. The final reason is that confirmation of Locard's theory was experimentally proven in the last 30 years of the twentieth century by using fibers as transfer objects. In fact, many other types of particles may be transferred in a contact; thus the forensic microscopist who studies dust must have developed the ability to identify a wide range of particles by their microscopic characteristics. As an example, the discovery and analysis of minute, green spheres of spray paint, which had been sprayed on cotton fibers to color the headliner of a car, were the most significant piece of physical evidence linking the suspect to the crime scene in the oldest murder retrial in US history.

2. *Can the origin of a dust sample be discovered, or its location described, from its composition*? The answer to this question is often of practical significance in an investigation. Whether or not it can be answered depends on both the nature of the dust involved and the skill and experience of the scientist conducting the analysis. Dust, by its very nature, always contains clues to its origin. These clues may not always be found, their significance may be obscure or they may not have importance to the case at hand, but certain generalizations may be made in almost every case. Thus indoor dusts normally contain large numbers of fibers. If the source is a home, these fibers will be of textile origin and include fibers from clothing, carpets, curtains, upholstery, etc. Human hairs, as

well as animal hairs from pets and feathers from bedding or pet birds, are also commonly found. The composition of the dust will also be controlled by the particular room from which the dust originates. Thus, the bath and bedroom will contribute large numbers of skin cells, pubic hairs and particles originating from talcum powder, antiperspirants, cosmetics, soap, etc. Dust from the kitchen contains particles of spices, flour, starch, etc.

Homes are not the only type of indoor location. The dust from an office may contain many of the same particles found in the living areas of a home. However, the fibrous component of an office dust tends to be dominated by the presence of paper fibers rather than those from textiles. Dusts of outdoor origin, on the other hand, consist primarily of silt-sized mineral grains, pollen, spores, plant cells and tissue, leaf hairs, soot and charred particles from plant burning. The exact composition of an outdoor dust depends on its geographic source, whether it is urban or rural, the time of year it was collected and the types of industries at or near the source. It is possible to narrow the range of inquiry by specifically identifying components of the dust and then studying their peculiarities. For example, it is possible to determine if quartz sand originated from a dune in a desert or from a seashore, on the basis of the shape and surface texture of the individual quartz grains comprising the sand. The identification of the mineral grains as quartz would be made using the petrographic microscope; then the surface of these grains would be examined with the scanning electron microscope. If the sample were large enough to isolate some of the trace (accessory or heavy) minerals, it would be possible to draw further conclusions regarding its geological origin once they had been identified with the petrographic microscope. Thus, although beach sands from the east and west coasts of the United States look similar at first glance, their accessory minerals are different: those from the west originating from the relatively young Rocky Mountains and coast ranges; and those from the east, especially the southeast, originating from the old metamorphic rocks of the Appalachians. If our sample also contained pollen grains which, after identification, would give us an idea of the flora in the region, it should be possible to narrow the possible sources of the sand to a relatively small geographic region. Additional particles present in the dust may make it possible to describe the source more closely still. While this example shows the possible lengths to which such an examination can be taken, it is important to realize that the true origin may be complicated beyond what a sound scientific analysis can provide. Thus, the most recent source of this hypothetical sand might be from a pile at a construction site and not the beach where it was formed. Fortunately, sands used for civil engineering projects are not usually transported over great distances and the conclusions would still generally be correct. Still, this illustrates the pitfalls to which conclusions drawn from this type of dust analysis are subject and points out the need for caution in the interpretation of the results.

3. *Is it possible to determine a person's occupation from an examination of the dust on clothing?* This was one of the principal concerns of Locard's original studies and he reported great success in this area. While it is undoubtedly true that people do pick up dusts from their place of work according to the exchange principle, the nature of these dusts may not be particularly distinctive and, when they are, they may be deposited on a set of clothes or uniform which the worker only wears when on the job. Thus, office workers carry paper fibers on their clothing, but so do people working in other offices, and the style and quality of the clothing itself may indicate this without the need to analyze the dust. Occupations that produce or expose workers to distinctive particulate matter still exist, and in these cases characteristic particles may be carried to vehicles and back to homes if the subjects do not change and wash at their workplace. In a recent specific example of this, the home of the murder victim contained significant numbers of metal turnings, which were brought into the home by her son, a machinist, who stopped off frequently to visit after work.

Collection and Recovery of Dust Traces

The collection of dust traces at a crime scene must be carried out with great care, as the exchange principle applies equally to criminals, victims and evidence technicians. The most effective trace evidence laboratories either collect their own evidence or train specialists to collect evidence in an efficient and scientific manner, while at the same time taking precautions to prevent contamination as the evidence is being collected and packaged. These precautions extend to wearing a white cleanroom garment, hat and booties over the clothing when collecting evidence from the crime scene. When multiple sites or vehicles are being investigated, the evidence collector must change to a fresh set of barrier clothing to prevent possible contamination from occurring between crime scenes. Failure to take such precautions has caused legitimate concern about contamination to be raised in cases where positive trace evidence comparisons were made because evidence was carelessly collected. For example, in a bank robbery where two vehicles were used in succession during the getaway, evidence was

collected by investigators who visited and collected evidence from both of the vans in the same day while wearing the same normal street clothing. It is impossible to state categorically that contamination could not have occurred under these circumstances.

It is possible, in some cases, to seek out suspected traces of dust from specific locations or from items of clothing or vehicles by direct examination. In other instances it is necessary to collect the dust first and then sort through it later to search for fibers or particles that may have a bearing on the case. In a recent homicide, for example, the murderer removed iron grillwork from a brick wall to enter the victim's home through the window. Microscopical examination of the suspect's clothing revealed particles of apparent red brick dust. Subsequent analysis of these particles proved that they were brick and that they were, both in color and petrography, the same as the bricks from the wall from which the grill had been removed. In this case the facts that the suspect's clothing contained brick dust and that his occupation did not involve working with brick illustrated the truth of Louis Pasteur's statement that 'chance favors the prepared mind'.

There are several methods of collecting dust for analysis. Perhaps the simplest and most direct is *hand-picking*, with the aid of a stereomicroscope or loupe, using forceps or needles. This was the way in which the suspected brick particles were removed from the clothing in the previous example. The only other necessary aids are a source of bright light and steady hands. Hairs, fibers or particles recovered in this manner should be immediately placed on glass slides, in well slides or small Petri dishes, depending on the size of the item, and covered. The location and type of particle as well as the date and the location from which it was removed should all be noted. If at all practical, the specimen should be photographed *in situ* before it is isolated for analysis.

Other methods of collection are available for recovering dusts of smaller particle size than those that can be easily manipulated with forceps. Selection of one of these methods over another will depend on the item or location from which the particles are to be recovered and the purpose for which the dust is being collected. If, for example, the substrate is an item of clothing and it is being searched for evidence of recent contact, *taping* the surface with transparent adhesive tape is the method of choice. Experiments have shown that this is the most efficient method for recovering surficial dusts. This technique was developed by a student of Locard, the late Dr Max Frei-Sulzer of Zurich. It has the advantage that the tapes can be placed sticky-side down on clean slides or acetate sheets, thus preventing contamination while at the same time providing a clear view of the collected dust when searching with either a stereomicroscope or automated fiber finder. The sticky tape method is also of use when a specific but relatively small area is to be searched for trace evidence. Thus, the area around a point of entry into a building can be sampled efficiently by pressing tape around a window frame, broken glass, torn screen, etc. to collect any traces from the clothing of the person who passed through the opening. Occupational and environmental dusts which have penetrated the weave of clothing over time, and large areas such as the interior of a vehicle, may require the use of a *vacuum cleaner.* Other dust collection methods such as *washing* and *scraping* may occasionally be of value. Washing is useful when the particles are in a sticky or greasy matrix. Locard described a case in which he washed the hair of suspected bomb makers in a gallon of ethanol. When the alcohol evaporated, it left a thick residue of nitrocellulose, the principal component of smokeless powder. Scraping is rarely, if ever, used in European laboratories but is still in use in some laboratories in the United States. The item is hung over a large sheet of fresh paper and a large metal spatula is scraped over the surface. The dust that falls down on to the paper is collected and placed in small containers for later sorting under a stereomicroscope.

Analysis of Dust Traces

The complete analysis of a sample of dust can be a time-consuming and difficult exercise in microanalysis, but fortunately a full dust analysis is rarely required for forensic purposes. It is necessary at the outset of the analysis, however, to formulate clear and concise questions that the analysis is expected to answer. Thus, if the question were 'Are there any blue-dyed rabbit hairs from the victim's sweater on the suspect's clothing or in his car?', it would be necessary to first characterize the fibers from the sweater so that similar fibers might be recognized in the tapings from the suspect's clothing and vehicle. Such fibers are called target fibers and they give the microscopist something to look for in trying to answer the question.

Identification of the components of a dust presents special challenges because of the small size of the individual particles. While some substances can be recognized at sight under the relatively low magnification of a stereomicroscope, most microscopic particles must be isolated and mounted between a slide and coverslip in a specially chosen mounting medium. Particle manipulation is performed under a stereomicroscope using finely sharpened forceps and specially sharpened tungsten or steel needles, which have

points much finer than those used in common dissecting needles. Once mounted in a familiar mounting medium, many particles can be recognized, at least to their class, by a well-trained and experienced forensic microscopist. The ability to identify a large number of particles on sight or by using a few simple optical tests is an important skill and useful advantage for the trace evidence analyst. It means that considerable time can be saved and the microscopist can quickly focus in on the particles that may be most helpful in answering the questions that have been posed. Even the most experienced and skilled worker will still require the aid of other techniques and instruments to confirm an identity and, if necessary, perform comparisons of particles from dust. Some of the most useful techniques and instruments for the analysis of the components are listed and described briefly below.

Microscopy

Many particles can best be identified on the basis of certain morphological characteristics. Substances of this type usually belong to a group in which all the members are chemically similar. An example of this is cellulose. Wood and wood fibers, the nonwoody vegetable fibers, rayon and lyocell fibers all have nearly the same chemical composition, yet it is a simple matter to distinguish each of these groups from one another and within a group to identify the individual members on the basis of their microscopical characteristics. Botanical fragments from leaves, seeds, bark, flowers, stems and roots are all identifiable, with suitable training and reference materials, and in the same way are pollen grains. No amount of chemical analysis would identify these substances as accurately as a well-performed microscopic analysis. As for comparison, it is a general rule that the best comparison of such materials is an accurate identification.

The optical properties of amorphous or crystalline transparent particles and manmade fibers, measured or estimated with the aid of the polarizing microscope, extend the range of application of the microscope to the identification of materials without regard to their morphology. Thus glass, minerals, drugs and chemical crystals of all types can be identified, when encountered as single small particles, on the basis of their optical properties. The technique, when applied by a skilled microscopist, has the same advantages as X-ray diffraction, in that it is possible to identify a particle as being a particular phase and not just a particular compound. One could distinguish, for example, between gypsum (calcium sulfate dihydrate), anhydrite (calcium sulfate) and plaster of Paris (calcium sulfate hemihydrate) on the basis of the optical properties measured on a single small crystal. A classical chemical analysis of each of these compounds would show only calcium and sulfate ions, while X-ray spectroscopy would show only calcium and sulfur. Polarized light microscopy is particularly useful for rapidly identifying manmade fibers as a prelude to comparison.

Microchemical analysis

It is frequently necessary to obtain chemical information from small particles to confirm their identity, or to identify metallic particles or unusual organic or inorganic compounds that cannot be identified with certainty by microscopy. Classical microchemistry is a useful skill, as microchemical testing can be performed on inorganic, organic or metals with equal ease and speed on very small particles. Although microchemical testing with reagents has been largely supplanted by instrumental methods, such tests, when applied by a scientist trained in their use and application, are rapid and provide chemical information that may be difficult or impossible to obtain with other techniques. The detection of acid radicals or organic anions can be performed with certainty, the oxidation state of an element can be identified, and tests are available for classes of compounds (e.g. carbohydrates), to name just a few of the many possibilities. The principal types of microchemical reactions are based on crystal tests, spot tests and histochemical reactions that have been scaled down so that they can be applied to single small particles or small quantities of material. Crystal tests are usually performed on a microscope slide. The reaction products are crystalline compounds with characteristic crystal form, habit or optical properties. Spot-test reactions result in reaction products that are highly colored, gaseous or fluorescent. Positive histochemical tests usually result in the formation of highly colored or strongly fluorescent reaction products which form *in situ* on the particle itself. The value of microchemical testing in dust analysis is that it can rapidly provide chemical information on extremely small particles with very little cost in money or time.

X-ray spectroscopy

This technique is capable of providing an elemental analysis of extremely small particles and is thus well suited to the analysis of dust particles. The spectrometer is attached to a scanning electron microscope. The electron beam of the instrument generates X-rays from the specimen in addition to the imaging secondary and backscattered electrons. The energy of the X-rays are related to the atomic numbers of the elements

in the sample. In modern instruments, all of the elements, from boron up, are detectable. The advantage of this technique is that the analysis is simultaneous and the entire elemental composition of an extremely small particle (down to about 0.1%) can be obtained in seconds. It makes no difference if the sample is inorganic or organic, so both metals and paints, for example, can be analyzed with equal advantage.

Infrared microspectrophotometry

Just as the elemental composition of a small particle can be obtained from an X-ray spectrometer, its molecular composition can be determined by this instrument. It consists of a Fourier transform infrared spectrophotometer attached to a reflecting microscope. An infrared spectrum permits the analyst to identify organic compounds and anions by comparison to reference spectra or by interpretation of the infrared absorption bands in the spectrum. The microspectrophotometer makes it possible to collect the spectrum from particles as small as a few micrometers in diameter. Although there are many other sensitive instruments in the modern analytical laboratory, these have not, as yet, been employed to any great extent in dust analysis.

Reference Collections

Reference collections are essential for the forensic science laboratory that conducts dust analyses. These collections are composed of physical specimens of the small particles which comprise the world in miniature. The number of specimens and completeness in any category will vary by laboratory. In some areas the collections are built over time; in others they will have been acquired by purchase. The items to be included will depend on the availability and the interests of the analysts in the laboratory. Potential items for inclusion are too numerous to list completely, but a useful collection might include the following: metals (both different metals and alloys as well as metal particles which have drilled, sawed, crushed with a sledgehammer, etc.), industrial dusts, combustion products, paper (woody) and nonwoody vegetable fibers, manmade fibers, human and animal hairs, minerals, sands (from different geological environments), wood, pollen, spores, diatoms, leaves, seeds, food products (e.g. spices, grains, starches, spray-dried products), polymers, synthetic drugs, explosives, dusts from different environments (these can be collected from air filters in hotel rooms or by vacuuming particles from the clothing after returning from a trip), and occupational dusts (obtained by vacuuming the clothing of volunteers working in various jobs involving manual labor). Depending on the nature and the amount of sample available, items may be stored in small vials, envelopes or as permanent mounts on microscope slides. An entire collection may thus be stored in a several drawers in a few relatively small cabinets.

See also: **Analytical Techniques**: Microscopy; Spectroscopy: Basic Principles; Spectroscopic Techniques; Presumptive Chemical Tests. **Fibres**: Types; Identification and Comparison. **Hair**: Identification of Human and Animal Hair, Comparison: Microscopic.

Further Reading

Benedetti-Pichler AA (1964) *Identification of Materials*. New York: Springer.

Demmelmeyer H and Adam J (1995) Forensic investigation of soil and vegetable materials. *Forensic Science Review* December: 119–142.

Frei-Sulzer M (1949) Mikroskopische Analyse von Verbrecherspuren. *Mikroskopie* **IV** 11/12, Part I: pp. 176–181; Part II: 355–360.

Frei-Sulzer M (1951) Die Sicherung von Mikrospures met Klebband. *Kriminalistik* October.

Hamilton EM and Jarvis WD (1962) *The Identification of Atmospheric Dust by Use of the Microscope*. London: Central Electricity Generating Board.

Kirk PL (1960) *Crime Investigation: Physical Evidence and Police Laboratory*. New York: Interscience.

Krinsley DH and Doornkamp JC (1973) *Atlas of Quartz Sand Surface Textures*. Cambridge: Cambridge University Press.

Locard E (1930) The analysis of dust traces. *American Journal of Police Science*. Part 1: May–June pp. 276–298; Part II: July–August pp. 401–418; Part III: September–October pp. 496–514.

Mange MA and Maurer HFW (1992) *Heavy Minerals in Colour*. London: Chapman and Hall.

Nickolls LC (1956) *The Scientific Investigation of Crime*. London: Butterworth.

Palenik SJ (1979) The determination of the geographical origin of dust. In: McCrone WC and Delly JG (eds) *The Particle Atlas*, 2nd edn, vol. 1, pp. 1347–1361. Ann Arbor: Ann Arbor Science.

Palenik S (1997) Forensic microscopy. *Microscopy and Microanalysis*, 7–9.

Winchell AN (1954) *The Optical Properties of Organic Compounds*. New York: Academic Press.

Winchell AN (1965) *The Microscopical Characters of Artificial Inorganic Solid Substances*. New York: Academic Press.

E

EAR PRINTS

C van der Lugt, Institute for Criminal Investigation and Crime Science, Zutphen, The Netherlands

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0497

Historic Overview

It is only during the last few years that ears and ear prints, and the identification of (unknown) ear prints found at a crime scene and known ear prints from a suspect, have attracted the attention of the police and prosecuting attorneys. Nevertheless, the use of ears and ear prints as a means of establishing identity is quite old.

In olden days, for instance, the length of the ear lobe was considered to be a sign of great wisdom in some countries in Indo-China, which is why statues of Bhuddha always have long lobes. Aristotle considered long ear lobes to be a sign of a good memory. During the Renaissance, when physiognomy was introduced and the followers of this doctrine pointed out that the face is a reflection of all intelligent qualities of a human being, much attention was paid to the shape of the ear. During his research into primates, Darwin also attracted the attention of the scientific world by saying that the ear is one of the elementary organs. In support of this, he pointed to the broadening of the middle of the helix, indicating that this is nothing else but a corner of the primitive ear that has become reduced in size. Science acknowledged this by naming it the 'tubercle of Darwin'.

Schwalbe was one of the first to invent a method of measuring the external ear. He was able to prove Darwin's theory and was the first to attract scientific attention to racial peculiarities in the structure of the ear.

Quotes from others who expressed their feelings about the importance of the ear can be found in *Die Bedeutung der Ohrmüschel für die Feststellung der Identität* (The Meaning of the Auricle for Identification Purposes) by Imhofer of Prague. He quotes Lavater (1741–1801), who, in publications during the period 1775–1778, argues for the position of the ear to be considered as part of physiognomy. Imhofer also quotes Amédé Joux, saying: 'Show me your ear and I'll tell you who you are, where you come from and where you're going.' The author refers to his conviction that there is no other organ of the human body that can better demonstrate the relation between a father and his child. It was his belief that the shape of the ear is clear proof of paternity. The article continues by stating that ears can be very important in establishing identity in:

- the determination of the identity of corpses;
- the determination of the identity of living persons; for example, when establishing the identity of a wanted person, or establishing family relationships.

Imhofer quotes Bertillon by writing that: 'It is almost impossible to find two ears which are equal in all parts and which have many shapes and characteristics which stay without noticeable changes throughout a life time.' The article continues with a discussion of methods for comparing ears and the characteristic features necessary to determine identity, as well as the numbers of characteristic aural features found in comparative research, with 500 pictures of ears. A combination of three features only appeared in two cases and it was concluded that a combination of four features is enough to establish identity.

In the development of a system to protect parents, children and hospital administrations from cases of mistaken identity in babies, Fields and colleagues suggested using the appearance of the ears in newborn babies. Footprints, palmprints and fingerprints, in conjunction with some additional identification, were generally used in most hospitals. The problem with these prints was that there was no or little reference value for identification and the prints were often of such poor quality as to render them useless, even within the first month. After a test study, taking pictures of infants, it was noticed that the appearance of the ear remained constant, whereas other features of the face changed. When a questionnaire was sent to plastic surgeons throughout the English-speaking

world, 75% of the responses favored the use of the ear as a constant for identification.

A study was conducted and included photographs of the right and left ears of a series of infants, taken daily from the day of birth to the day of discharge from hospital. These photographs were taken to document the minute changes that took place in the growing ear during each 24 h period and to document the gross morphologic changes that took place from the date of birth to the date of departure from the hospital. The result of the research was that 206 sets of ears were photographed and the following conclusions were reached:

- The ears of different sample babies are always unique; no baby was found to have ears identical in size or configuration to those of another.
- There is enough variation in individual ear form to distinguish visually the ears of one baby from those of another.
- The sample series of the photographed ears of a baby remained homogeneous (markedly similar in form and structure) during the entire hospitalization period.
- The minute changes that take place in the growing ear are so slight that the immutability of the series is not jeopardized.

The described photographic research procedure was deemed to have met all of the requirements of a reliable and standardized identification technique: individuality, continuity and immutability.

The identification of individuals by ear characteristics has been described in a text for scene-of-crime officers by Söderman and O'Connell. The principal parts of the ears are described, as well as peculiarities of the helix and the antitragus, descriptions that were said to be useful for identification purposes. An ear print case was presented by Medlin, giving a description of the way the case was handled and the results. The actual procedure of the comparison is not explained.

The characteristics of the auricle, measurements of the ear and morphological characteristics with genetic importance have been studied by Oliver, who described variations in the form of the ear, the darwinian tubercle and lobular adherence.

Oepen presented the morphological aspects of ears, in a lecture and articles for the German Forensic Medicine Society, from research on the differences between the features of the ears and their appearance in practice in 500 men and women. The study was done to find out what features of children's ears were characteristic and could indicate the identity of the biological father. The outcome of the research was a useful contribution to the comparison of ears. Several tables illustrate the findings regarding aural features and give statistical information about how often a combination of features was found. Other tables show the features that were found more often in women then in men, and vice versa, and more often on the right side than on the left side, and vice versa. Attention is paid to the way in which an ear will print and what features of the ear will be visible.

Hirschi from Bern in Switzerland dealt with a burglary case in which ear prints were found. The ear prints of a left and right ear could be identified as placed by suspect 'P'. Besides the information as to on what ground the prints were identified, the report also contains quotes from the well-known and respected scientists Reiss, Niceforo and Locard regarding the value of ears in the identification process. Hirschi also gives reference prints from test persons that were used for the comparison. He describes the work that had been done regarding the height at which the ear print was found and the height of the perpetrator. Hirschi carried out research among 40 recruits from the Swiss Police Academy as to the degree to which they bent over; the distance between the top of the skull and the middle of the auditory canal; and whether they were right-handed or left-handed.

Hunger and Leopold discussed medical and anthropological views regarding the identification of people and referred to the shape of the ear being basically specified by its cartilage. They pointed out that the external ear has many specific features useful for identification, particularly because these features are very stable throughout the aging process and even after death. They note that not all of the characteristic features may be recognized on photographs, which also applies to prints.

Trube-Becker from Dusseldorf pointed out that there are no absolutely identical ears, but only similar ears. Even two ears of one and the same individual are not completely identical. This is equally true for identical twins. A warning is given with respect to the fact that the properties of ears can change as they are pressed against a wall or a door. Nevertheless, as they are clearly recognizable, comparison is possible. The opinion is stated that ear prints can at least provide additional circumstantial evidence for a suspect's guilt and that it is therefore worthwhile looking for ear prints and preserving them like fingerprints. They can provide an additional piece in the mosaic of identification of the perpetrator.

Georg and Lange described the case of a bank robbery in which a comparison was made of several features of the human body for identification purposes. They expressed the view that multiple morphologic variations (not only in facial characteristics but also differences in, for example, the features of

the hands and fingernails, height and body shape) are at least of the same value in identification as fingerprinting. The ear in particular provides many morphologic and metric features: as with fingerprints, no two people in the world will have ears that have the same combination of features. On the basis of several morphologic features in this particular case, a suspect was apprehended. Statistical calculations showed that the combination of morphologic features in this case could only be found 1 in 300 000 000 people.

Cases involving ear prints in the surroundings of Dusseldorf have been investigated by Trube-Becker. She renders an opinion on several crime scene prints and states that many of these prints originated from a suspected perpetrator 'with a probability bordering on certainty'. She describes the work that was done by the Ludwigshafen Criminal Investigative Police, who apparently found ear prints repeatedly. The prints were rendered visible with soot dust or a Magna Bush and lifted with foil. For making the standards (controls) from suspects they developed a system using a Polaroid CU-5 camera with a transparent macrolon disk on its front. They were able to place the disk against the head and actually see the change in the form of the ear caused by the pressure. The prints were then made visible with fingerprint powder and lifted with foil. At the same time the ear was photographed at a scale of 1:1. A built-in ring flash allowed a colored picture to be taken but it appears that the flash is always visible on the photograph, which has such a damaging effect on the result that the photograph cannot be used for the purpose for which it was intended.

There have been cases in Freiburg (Germany) with findings of ear prints in a series of burglaries. Meyer relates a case in which two girls were seen near the site of the burglaries. Police suspected them of having listened at the door before breaking in. The girls were arrested for identification purposes, during which prints of the left ear were taken. The next day a message was received from Stuttgart, where three girls had been arrested for the same type of burglary. Prints of these girls' left ears were also taken. Investigations quickly showed that (according to the ear prints) the girls from Stuttgart could not be excluded from having committed the burglaries in Freiburg. The prints from the scenes of crime were taken to a fingerprint expert in Baden-Württemberg, who stated that: 'The traces found at the different crime scenes were identical and caused by a left ear. The prints contained more or less the same part of the ear corresponding to the cartilage of the ear. The traces of the ear prints contained each at least 8 of the 12 most important points of the ear that can be notified.' A comparison was made between the prints from the crime scenes and the standards from the girls. It was found that the prints were identical to the prints from one of the girls arrested in Stuttgart. The expert stated: 'Between the standard of the left ear and the three prints from the crime scene there is, taking into account a tolerance defect, the highest match in form of coverage equality'.

The anatomical design of the auricle is so individual and multiple that, with a probability bordering on certainty, it can be stated that no two ears can be found that are identical in all details. Hammer summarizes the value of ear print identification by saying that, of 100 ears investigated, no two ears could be found to correspond in all features and that the human ear print is therefore suitable for establishing a person's identity in forensic practice, although the reliability of the information it gives on questions of identity depends on the quality of the ear prints secured at the scene of the crime.

Hunger and Hammer, from many years' practical experience, stated that the ear can play an important role in the identification of both perpetrators and unidentified corpses. They describe the results of research into the distribution of selected ear characteristics in the East German population, recording which types of aural features were found to be rare in a group of 350 men and 300 women. The authors also made two concluding observations. First, as regards the identification of the ear from a severely dismembered corpse, special care must be taken in preparing the ear for preservation by photography. It appears that characteristics can be altered when the tissues in the head area are separated from their surroundings. The second point concerns ear prints, of which it is said that when the ear is pressed firmly against a surface, the ear can change, particularly in size and shape. Hunger and Hammer state that the metric and morphologic characteristics of the ear are very well suited for determining identity, especially when rare characteristics are present.

The first case in which ear prints led to a conviction by a Dutch court of law, reported by Dubois, concerned a burglary and hostage-taking. The scene of crime investigation and the procedure leading to the ear identification are described, together with the examination of 100 test subjects to find out whether or not their ears were similar. This research showed that the ear of none of the test persons resembled the ear found at the scene. In fact, great dissimilarities were found, whereas the ears of a suspect were identical. The brothers of the suspect were asked to contribute to the investigation to find out whether the suspect's ears were similar to his brothers' ears. The results were very much to the contrary. During the investigation a forensic odontologist and an ear, nose

and throat specialist were consulted. The court at Dordrecht accepted the ear prints as evidence and convicted the suspect. The court of appeal in The Hague only used the ear print as supporting evidence and convicted the suspect, partly on the basis of this but mostly on other evidence in the case.

Rochaix has said: 'The ears are after the fingerprints the best means for identifying people, because they stay the same from birth till death. They do not change except for the size. The ear is the most characteristic feature of the human being'. He also describes the form and features of the human ear and points to the work done by Bertillon, Reiss and Locard, as well as the method used by the cantonal police at Vaud (Switzerland) for recovering and lifting prints and taking the standards from a suspect. He introduces a classification method for ear prints based upon several features of the ears, which will usually be visible after listening. In total there are five features to be used for obtaining a final classification code; these items include the shape of the ear; the attachment of the ear lobe to the cheek; the bending and basis of the antitragus. These codes give a series of numbers for both left and right ears and can therefore be stored and found easily. Rochaix concludes that, using this classification, 600 different ear prints can be stored separately. Whether or not this system is adequate to separate the different ear prints found needs verification.

Iannarelli, in 1989, introduced a classification system for ears in the United States.

Hammer and Neubert summarized their research and that of Hirschi with respect to measurements of height. They found that the distance between the top of the skull and the middle of the auditory canal was 13.1 cm for men and 12.4 cm for women in the East German population. Similar values were found in a study in 1986, namely 13.2 cm for men and 12.5 cm for women. They conclude that it is only possible to ascertain the height approximately because: (1) the middle of the external ear cannot always be determined; (2) the inclination of the head while listening averages 3 cm (according to Hirschi) but the range is 1–8 cm; (3) although average ear height might be known, there is about 4 cm between the minimal and maximal values. The authors also give the results of research into pressure distortion. The differences between ears that were pressed weakly and strongly against a surface were investigated at eight different points of the ear and differences in four descriptive ear characteristics were noted. Among the 100 ears that were analyzed, not one of them was completely identical to another.

Kritscher and colleagues described the identification of a bank robber partly based upon comparison of the ear. The available information came from a surveillance camera, which showed a man of medium height with only the right ear clearly discernible and thus useful for identification purposes. They describe the method and measurements used for comparison. The images of the bank robber's right ear, as caught by the surveillance camera, and the suspect's right ear proved identical in virtually all respects. The authors concluded by saying that their morphologic–anthropometric examinations enabled them to ascertain the identity of the bank robber to a large degree, in particular as the matching morphology of the auricles could be rated as strong a proof as are identical fingerprints.

There are those who doubt the possibility of identifying a person by their ears or ear prints. Moenssens gives his basic reason for doubt by stating that there has been no empirical research on the subject. On the other hand he writes: 'A close look at the external part of the ear shows that it contains numerous complicated details which may be arrayed in particular combinations possibly unique to each person.' He continues by saying: 'Forensic anthropologists recognize the possible individuality of an individual's ear, but not as a means of identification via ear impressions.' Attention is given to the fact that there is a significant difference between comparing actual ears and photographs of ears, and comparing ear prints to each other; and it is concluded that a reliable and replicable standard procedure has yet to be developed. There have not been any professional conferences on the subject, nor any court decisions at the appellate level recognizing such a discipline and admitting evidence based upon its comparisons.

In Poland a study of 31 cases involving ear prints was carried out over a 5 year period. Labaj and Goscicki give advice on how to recover and lift ear prints and how to take into account the height at which the print is found. They use a distance of 17 cm to arrive at the estimated height of the perpetrator.

A literature survey and practical research on the influence of pressure on ear prints has been carried out at the National Training Centre at Durham, UK. Saddler has reported on ear prints, with an overview of the materials and the methods used to obtain them, an explanation of the method used for measuring and a discussion on the development of a computerized database.

The present author's own research on the relation between the height of an ear print at the scene of a crime and the height of the perpetrator was similar to the research done by Hirschi in Switzerland. It was carried out between 1990 and 1992 at the Dutch College for Criminal Investigation and Crime Control at Zutphen. The difference between the two

investigations related to the number of test subjects used in the research and the fact that the Dutch research studied both men and women and used a different method. The findings of this research were also different. The distance between the top of the skull and the middle of the auditory canal appeared to be 14 cm on average for men and women; the inclination was on average 6 cm.

In Rumania, ear prints are also used for identification. Pasescu and Tanislav give an overview of a number of cases handled since 1975. The number of cases increased after the publication in 1976, by the Institute for Forensic Studies, of a practical handbook of forensic investigation, which included a synthesis of the main scientific observations on ear prints that had been published up to that time. They describe the sites at which ear prints were found and the way in which they were recovered and lifted.

Although ear print evidence has been used in practice in the Netherlands since 1986 (the Dordrecht case), and more frequently since 1992, the attention paid to this type of evidence (although it has been included in scene-of-crime officer training since 1991) increased after a murder case in 1996 in England, followed by one in Washington State. The American case included a FRYE-hearing about the admissibility of ear prints as well as the trial itself. The case led to the first admission of ear print evidence in the United States, as well as the first conviction based upon this evidence.

Morphology of the Ear

Basically the ear consists of cartilage, which gives the original shape and dimensions to the ear, covered with skin. The development of the ear starts shortly after conception and by the 38th day some of the features of the ear are recognizable. The ear moves to its definitive position on about the 56th day and the shape of the ear can be recognized on the 70th day. The shape is normally fixed from then on and never changes from birth until death.

Figure 1 illustrates the external features of the ear. Ears can be divided into four basic shapes: oval, round, rectangular and triangular (**Fig. 2**). These shapes occur in all races but the percentage of each shape differs between races. The difference in the shape of ears established from ear prints is just a class characteristic.

For comparative purposes, the individual appearance of the features of the ear, their dimensions and their relation to other features are of most importance. A combination of features that can be identified can lead to the individualization of the ear print.

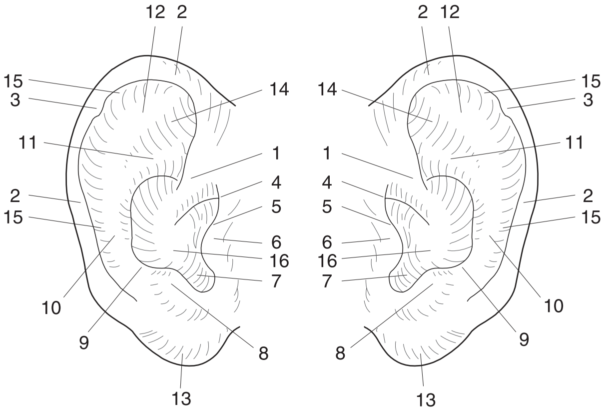


Figure 1 The external features of the ear: 1, crus of helix; 2, helix; 3, auricular tubercle (of Darwin); 4, anterior notch; 5, anterior tubercle; 6, tragus; 7, intertragic notch; 8, antitragus; 9, posterior auricular furrow; 10, anthelix; 11, lower crus of anthelix; 12, upper crus of anthelix; 13, lobule; 14 triangular fossa; 15, scaphoid fossa; 16, concha.

Where to Find Ear Prints

Experience shows that ear prints are found predominantly on surfaces where an individual has been listening to determine whether or not a premises was occupied. This occurs generally at doors or windows;

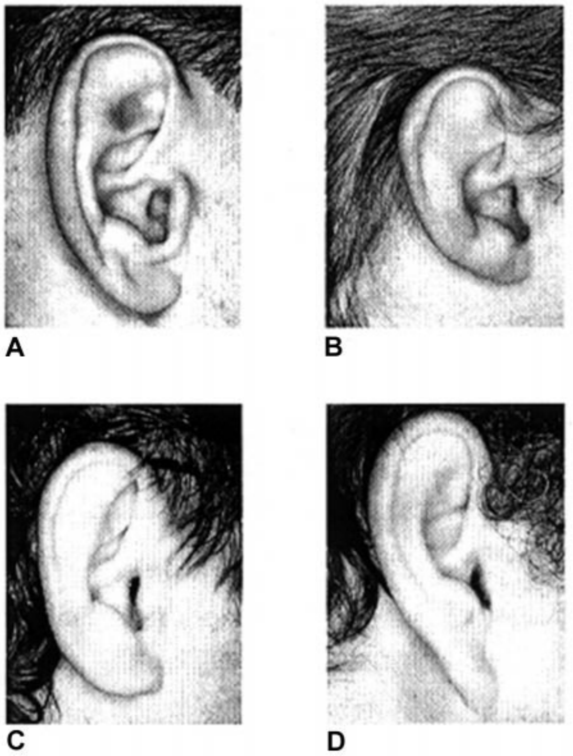


Figure 2 Ear shapes: (**A**) oval; (**B**) round; (**C**) rectangular; (**D**) triangular.

hence the surfaces that are likely to be examined are predominantly glass, wood or painted (**Fig. 3**). However, it may not be at the door or window where entry was gained, or even the same premises, that an ear print is found: individuals may have listened at several premises along a street, or several windows in one house.

In the Netherlands the majority of ear prints are found on doors in blocks of flats. They are found at two different heights: one close to the ground and one at normal standing height. Within any block of flats, ear prints may be found on doors and windows on one floor or several floors, not all premises having been entered.

The first procedure, which is always utilized when attempting to recover ear prints from any type of surface, is a thorough visual examination. It is essential that there is good lighting, either natural or artificial, and considerable care is necessary to carry out a thorough and effective examination. Artificial lights should ideally produce a high level of flat and even illumination, rather than the glare produced from some sources. Additionally, combinations of change, involving the angle of light, the viewing position and where possible the rotation of the article, will significantly affect the success rate. Some ear prints can only be seen clearly by the use of oblique light. Care is needed when handling articles using this technique so as not to damage other ear prints that may not be visible at this stage.

The latent impressions can be made visible by applying a suitable reagent, e.g. fingerprint powder. In the case of ear prints that are fresh, the powder principally adheres to the aqueous component of the sweat deposit, whereas with older ear prints fatty deposits become much more relevant.

Traditionally, the most widely used and productive method of recovering fingerprints is the application of fingerprint powder, followed by photography or lifting of the marks found. This method can also be readily applied to ear prints (**Fig. 4**).

Influence of Pressure

Because of the influence of varying pressure, ear prints recovered from a crime scene will never be identical in every respect to standards (or controls) taken from a suspect. This is due to the fact that the exact pressure exerted at the crime scene is an unknown quantity and therefore cannot be reproduced. Likewise the direction from which the pressure was applied cannot be duplicated exactly. Although it is

Figure 3 The manner in which someone listens at a door.

Figure 4 Application of fingerprint powder to an ear print.

impossible to recreate the exact pressure of the ear and rotation that left the impression, it is possible to create a similar impression so that pressure features and other characteristics may be compared. In fact a 'duplication' of the amount of pressure is not necessary.

In summary, unlike fingerprints, where identical marks are left, no ear print impression will ever be identical but it will have points of similarity that can be compared.

Depending on the applied pressure, features of the human ear may be visible, or not yet visible. By applying more pressure it is likely that more features of the ear become visible, depending on the individual characteristic features of the ear. Features that were already visible will become clearer when more pressure is applied. Through exertion of pressure the so-called 'pressure points' of the ear will become visible and may leave a characteristic mark. To be able to study the changes in the appearance of the ear by application of pressure, at least three different standards or control samples need to be taken from a suspect. These must illustrate distinct pressure variations. By studying the differences caused by pressure, one can see or predict how an individual ear will change with pressure.

In addition to pressure, the effect of rotation of the ear when listening at a door or window must be taken into account.

Producing Standards (Controls) from Suspects

Several different methods have been employed to produce standards from suspects in an attempt to identify perpetrators. Most documented methods for obtaining standards include photographing the ear, in addition to any other technique employed. If a suspect cooperates in the investigation, standards should be made as follows.

The suspect is asked to 'listen' three times at a glass pane in a door (**Fig. 5**). The pressure that is exerted on the pane must be different each time: the first print taken is from the application of gentle pressure; with normal pressure; and the third with much more pressure (**Fig. 6**).

This method of taking prints is used for both right and left ears. In all cases a photograph of both ears is taken, with the camera at an angle of 90° to the head (**Fig. 7**).

The prints thus obtained are made visible in the same way as are fingerprint identification traces, and lifted with black lifter.

For suspects who will not cooperate, and those who

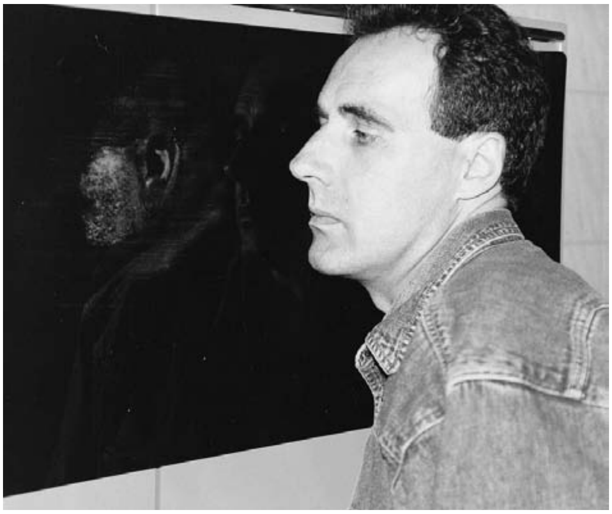


Figure 5 Taking standards (controls) from a cooperative subject.

offer passive resistance, standards are taken using glass or a synthetic plate (**Fig. 8**). Five prints of each ear must be made, putting variable pressure on the ear while preventing one part of the ear (upper, lower, front or back) having extra pressure placed upon it.

Procedures for Comparison

The procedure for comparison has two aims: first, to establish whether or not the ear print in question contains enough features for a comparison; and, second, to be able to tell the court that a strict procedure was followed every time an ear print was compared. This procedure consists of six steps.

Step 1 (analysis)

Note down every detail of the print in question:

(a) What type of fingerprint powder (or chemical) was used to reveal the latent print?
(b) What type of lifter was used to lift the print?
(c) What features of the ear are visible? (Use a picture with anatomic features to note down the details.)
(d) Which of the features might be specific for these ears?
(e) What pressure points are visible on the earprint?
(f) What additional information is visible? For instance: skin texture of cheek, head, etc.; hair; scars, pimples or birthmarks.

Is the information enough for comparison with a known print?

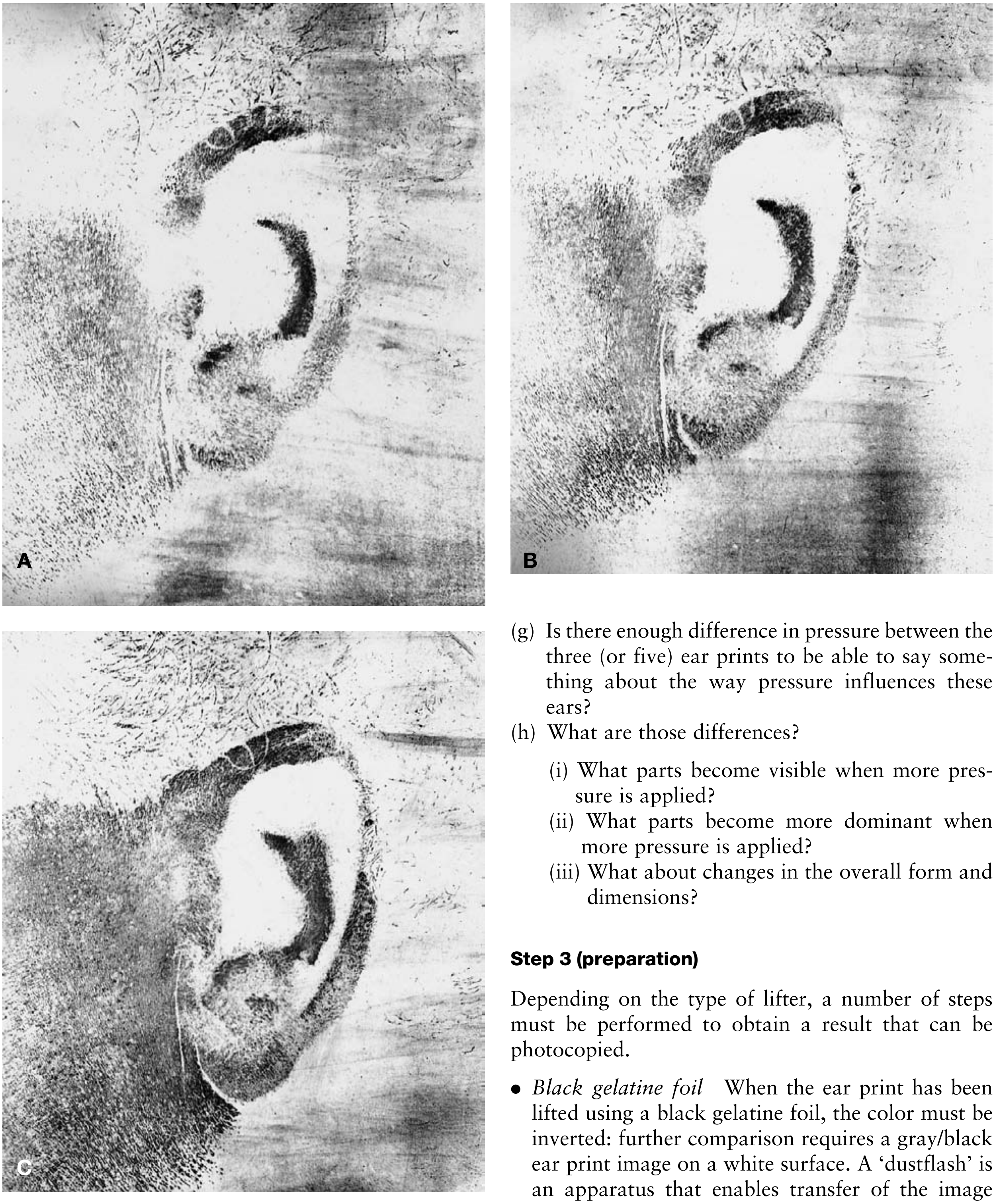


Figure 6 Prints left by (**A**) light pressure; (**B**) medium pressure; (**C**) heavy pressure.

Step 2 (analysis)

The same process applies to the known prints (the standards taken from a suspect). Follow step 1 (a)–(f); in addition:

(g) Is there enough difference in pressure between the three (or five) ear prints to be able to say something about the way pressure influences these ears?

(h) What are those differences?

- (i) What parts become visible when more pressure is applied?
- (ii) What parts become more dominant when more pressure is applied?
- (iii) What about changes in the overall form and dimensions?

Step 3 (preparation)

Depending on the type of lifter, a number of steps must be performed to obtain a result that can be photocopied.

- *Black gelatine foil* When the ear print has been lifted using a black gelatine foil, the color must be inverted: further comparison requires a gray/black ear print image on a white surface. A 'dustflash' is an apparatus that enables transfer of the image directly on to a photographic paper on a 1:1 basis. Photographic development will reveal the image.
- *White gelatine foil* This type of foil is not recommended for ear print lifting. The powder that is usually employed is a black powder (a soot powder) that does not adhere very well to the latent print. If this type of foil is used, the procedure to follow depends on the visibility of the print. Good

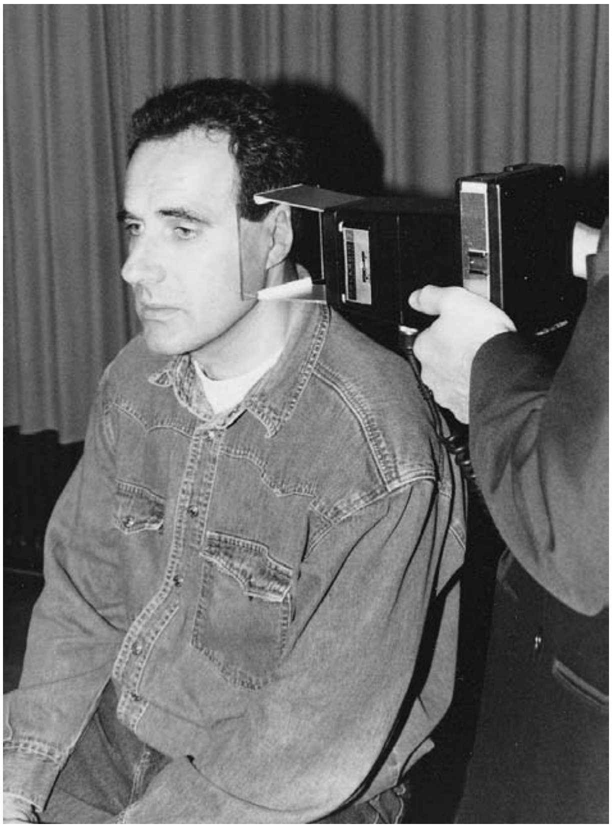


Figure 7 Photography of the ear.

prints can be photocopied directly; in other cases the print should be photographed and printed (on a 1:1 basis).

- *Transparent foil or tape* This type of foil (either used in combination with a gray aluminum powder or a black soot powder) can either be directly photocopied of photographed.

In all cases a digital image can be made, providing the print is reproduced on a 1:1 basis.

Step 4 (preparation)

Photocopies of both the known and unknown ear prints are produced on plain and transparency overlays. Be sure to have marked the traces well, or do this before making the copies. The unknown prints are copied on paper and the known prints on paper and transparency overlays. All of the prints must be reproduced on the same photocopier to ensure that any enlargement is uniform.

Step 5 (comparison)

The paper copy of the unknown print is taped to the top of a light box. The transparency overlay is put on top of the unknown print (**Fig. 9**). The procedure can be divided into the following steps:

(a) Put the tragus point (if visible) of the known print on top of the tragus point of the unknown print.
(b) Orientate the overlay sheet so that the crus of the

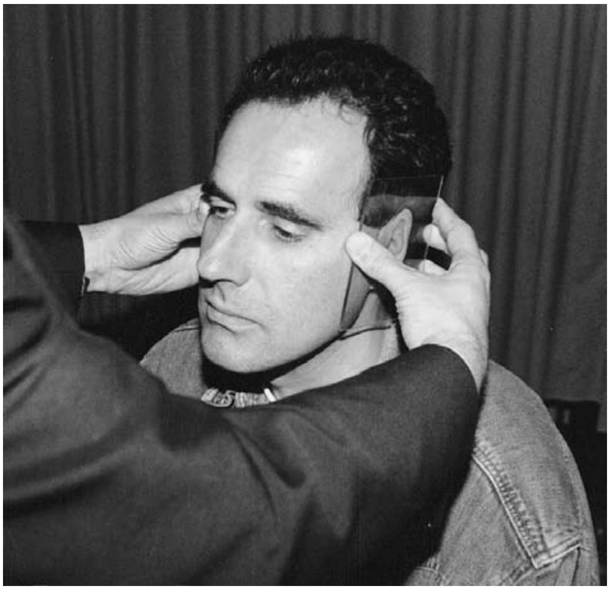


Figure 8 Taking standards (controls) from a subject using a synthetic plate.

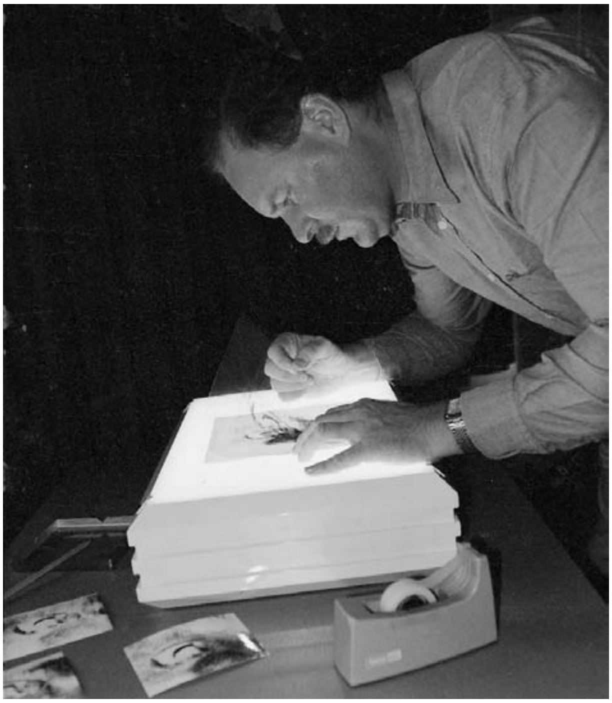


Figure 9 Using the overlay technique for comparison.

helix points of both unknown and known prints are in the same position.

(c) Look for the antitragus point and try to cover this point on the unknown with the known.
(d) If these three points more or less fit, look for the antihelix and see whether or not the shape and dimensions are alike.
(e) Look at the lower and upper crura of the antihelix (if present) and see whether or not they fit.
(f) Do the same for the helix rim.
(g) Look at any peculiarities of the helix rim (for instance a tubercle of Darwin) and see whether or not they fit.

If all details more or less coincide you may conclude that the prints are from the same source. If so, continue by:

(h) Attaching the transparency overlay to the photocopy (with a piece of tape).
(i) Cutting both copies into a square, or in case of a resemblance in the skin texture of the cheek, a rectangle.
(j) Attaching both to a piece of paper for presentation purposes.

Step 6 (comparison)

If step 5 was successful, repeat (a) to (h), using the two photocopies of the same trace. The next steps in this procedure are as follows:

(a) Take the ear template, using the stainless steel cross, and put it on top of the prints.
(b) Position the template so that as much as possible of the features of the prints of the ear are cut into two parts.
(c) Remove the stainless steel cross and mark off the inside of the template with a pencil.
(d) Cut the copies along the pencil line, using a roller cutter. A square (10 × 10 cm) will be produced.
(e) Divide the two copies (together) into four parts, according to the indication marks on the top copy (**Fig. 10**).
(f) Mark the different parts of the copies (known and unknown) carefully with, for instance, a letter and the numbers 1 through 4.

The different parts can be paired and attached to a piece of prepared 10 × 10 cm paper marked with two squares. The new combined images will consist of two different ear prints, but all of the features and dimensions will fit (**Fig. 11**).

All that remains is the evaluation, deciding in what terms you will express your opinion. It is difficult to give a standard type of report for the comparison of ears and ear prints. Local legal procedures and practices are usually the factors that influence its form and contents.

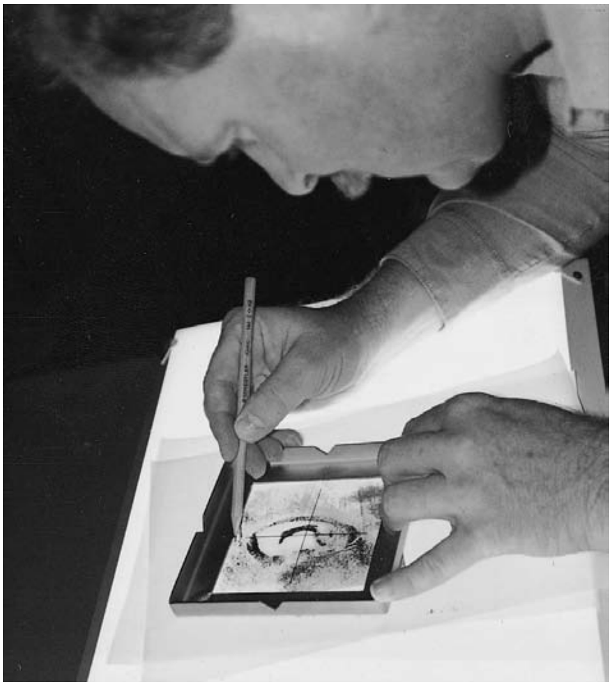


Figure 10 Using the template to divide the photocopy into four parts.

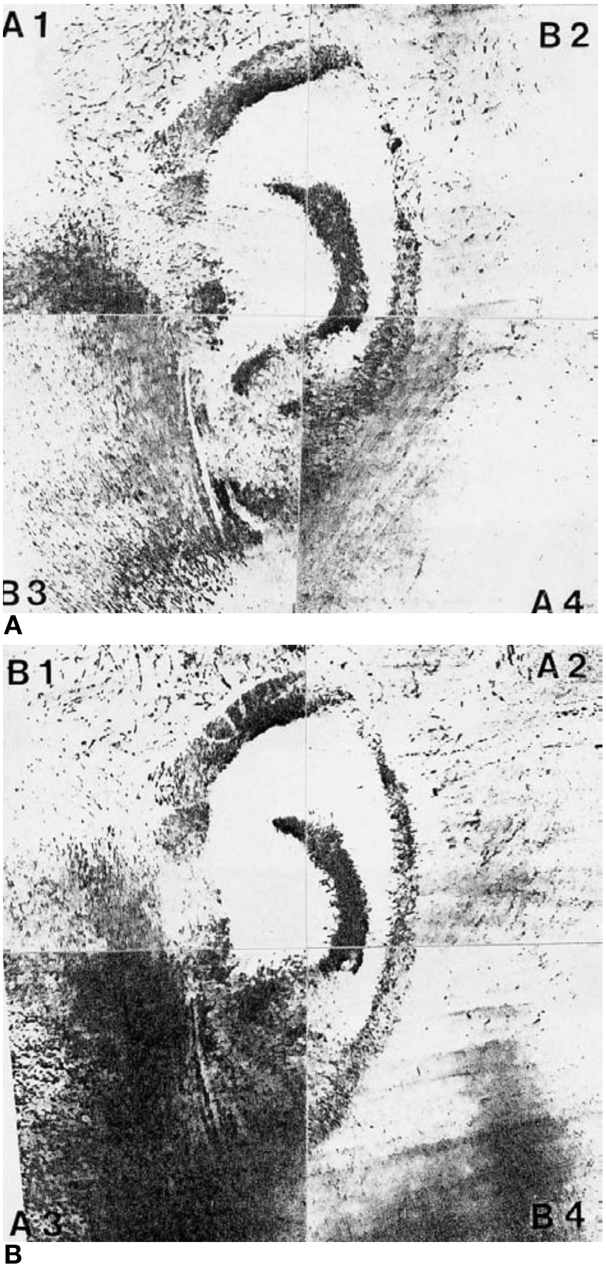


Figure 11 Combining the prints of the known and unknown ears to give (**A**) A1, B2, B3 and A4; and (**B**) B1, A2, A3 and B4.

See also: **Cheiloscopy**. **Fingerprints (Dactyloscopy)**: Visualization; Identification and Classification. **Identification/Individualization**: Overview and Meaning of ID.

Further Reading

Dubois N (1988) Oren naar hebben. *Technisch Informatie Bulletin* **1**:7–22.

Fields C, Falls HC, Warren CP and Zimberof M (1960) The ear of the newborn as an identification constant. *Obstetrics and Gynecology* **16**:98–102.

Georg K-H and Lange G (1983) Das anthropologische Vergleichsgutachten: eine sichere Methode zur Identifizierung von Straftatern. *Kriminalist* **15**:167, 169, 171.

Gerasimov MM (1949) The Ear. In: *Osnovy Vosstanovlieniia Litsapo Cherapu*, pp. 49–50.

Hammer H-J (1986) The identification of ear prints secured at the scene of the crime. *Fingerprint Whorld* **October**: 49–51.

Hammer H-J and Neubert F (1989) Experimental investigations into the evaluation of ear print traces. *Criminalistics and Forensic Science* **73/74**:136–139.

Händel K (1983) Streiflichter (Sidelights). *Kriminalistik* **1**:53–56.

Hirschi F (1970) Identifizierung von Ohrabdrucken. *Kriminalistik* **24**:75–79.

Hirschi F (1970) International burglars convicted on unusual evidence. *International Criminal Police Review* **239**:184–193.

Hunger H and Leopold D (1978) Merkmale der Ohren. In: *Identifikation, Medizinische und anthropologische Gesichtspunkte zur Identification Lebender: Korperliche Merkmale*, pp. 391–405.

Hunger H and Hammer H-J (1987) Concerning questions of identification by ear prints. *Criminalistics and Forensic Sciences* **65/66**:75–79.

Iannarelli AV (1989) *Ear Identification*. Forensic Identification Series. Paramount.

Imhofer R (1906) Die Bedeutung der Ohrmüschel für die Feststellung der Identität. *Archiv für Kriminologie* **26**:150–163.

Jung W (1984) Uber das Ohr zum Tater. Identifizierung durch Ohrabdrücke. *Kriminalistik* **10**:482–483.

Kritscher H, Szilvassy J and Bauer G (1994) Identifying a bank robber. *Acta Medicinae Legalis* **XLIV**.

Labaj E and Goscicki J (1995) The traces of the auricle in the practice of the capital's police criminalistics laboratory. *Problemy Kryminalistyki* **209**:40–45.

Medlin HO (1967) Earprint identification solves crime *Military Police Journal* **January**: 10–11.

Meyer K (1984) Identifizierung von Tätverdachtigen anhand von Ohrabdrücken *Kriminalistik* **12**:519.

Moenssens AA, Starrs JE, Henderson CE and Inbau FE (1995) Identification by ear impressions. In: *Scientific Evidence in Civil and Criminal Cases*, 9th edn. para 9.

Neubert (1985) *The Significance of Perpetrator Identification with Ear Prints*. MD thesis, Leipzig.: Karl Marx University, Leipzig.

Oepen I (1970) Uber die Häufigkeit einiger morphologischer Besonderheiten im Kopfbereich. *Anthropologischer Anzeiger* **September**: 141–156.

Oepen I (1976) Der Identifizierungswert des menschlichen Ohres. (The identification value of the human ear.) Paper presented at the Institut für Rechtsmedizin der Universität Marburg.

Olivier G (1969) The ear. In: *Practical Anthropology*, pp. 56–58.

Osterburg JW (1982) Earprints. In: *The Crime Laboratory:*

Case Studies of Scientific Criminal Investigation, sect 4.6.
Pasescu G and Tanislav E (1997) Person identification on the basis of earprints in the activity of Bucharest police department. *Information Bulletin for SP/TM Examiners* **3**:10–14.
Rochaix P-L (1988) *Classification des Traces d'Oreilles*. Institute de Police Scientifique et de Criminologies, University of Lausanne.
Schwalbe G (1897) Das Aussere Ohr. In: *Handbuch der Anatomie*, pp. 3–80.
Söderman H and O'Connell JJ Identification of individuals. In: *Modern Criminal Investigation*, 5th edn., chap 5. (Revised by O'Hara CE).
van der Lugt C (1977) Determining a person's height based upon the distance of a located earprint. *Journal of Forensic Identification* **47**:406–419

EDUCATION

An International Perspective

A Longhetti, California State University, Los Angeles, CA, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0498

Introduction

To qualify as a profession, forensic science must have peer-defined standards of competence, a code of ethics, a body of specialized knowledge and a strong educational core. Unfortunately, there are few forensic science programs, and very few that are considered to be grounded in a strong natural science background. Consequently, there are many nonforensic graduates working in crime laboratories. These individuals receive on-the-job training in order to acquire the necessary technical skills for their job and are often not able to acquire criminal justice philosophy and practice, and forensic science ethics, attitudes and perspective. Forensic science education can address these topics in its curriculum.

Along with the question of whether or not forensic science is a profession is the concern as to whether or not it is, in fact, a science. Practitioners, as expected, agree that it is both. Users of forensic science, and the general public, however, may not think so. Nonetheless, a forensic scientist's hallmark is an attempt to determine how events of legal significance occurred, based on observation, the formation of hypotheses and experimentation. This can easily be interpreted as a definition of science.

Unlike criminalistics, the other forensic sciences such as toxicology, pathology, anthropology and odontology have long-established education programs leading to an advanced doctoral degree. Consequently, this article is directed principally toward forensic science education as it applies to criminalistics (as defined and explained in detail in the article covering that field).

Education Surveys of Forensic Science Practitioners

Surveys of crime laboratory directors were carried out by Siegel in 1986 and Furton *et al.* in 1997. Others have written about the status of forensic science degree programs. Listings of such programs have been compiled by: (1) the Council of Forensic Science Educators (CFSE), an informal group organized under the auspices of the American Academy of Forensic Sciences (AAFS); (2) the United States Bureau of Alcohol, Tobacco and Firearms (BATF); and (3) the BC Institute of Technology (BCIT) in Canada. Some compilations of forensic science degree programs are available on the Internet.

Siegel's survey instrument was directed at three groups of respondents. One group, which comprised 23 forensic science majors at Michigan State University showed preference for: (1) a Bachelor of Science program in Forensic Science (BSFS) plus a Master of Science in Forensic Science (MSFS); or (2) a 5 year program leading to an MSFS. A second group consisted of 90 forensic scientists in the Michigan State Police, whose preference was an MSFS or a Master of Science in Physical or Biological Science. The third

Case Studies of Scientific Criminal Investigation, sect 4.6.

Pasescu G and Tanislav E (1997) Person identification on the basis of earprints in the activity of Bucharest police department. *Information Bulletin for SP/TM Examiners* **3**:10–14.

Rochaix P-L (1988) *Classification des Traces d'Oreilles*. Institute de Police Scientifique et de Criminologies, University of Lausanne.

Schwalbe G (1897) Das Aussere Ohr. In: *Handbuch der Anatomie*, pp. 3–80.

Söderman H and O'Connell JJ Identification of individuals. In: *Modern Criminal Investigation*, 5th edn., chap 5. (Revised by O'Hara CE).

van der Lugt C (1977) Determining a person's height based upon the distance of a located earprint. *Journal of Forensic Identification* **47**:406–419

EDUCATION

An International Perspective

A Longhetti, California State University, Los Angeles, CA, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0498

Introduction

To qualify as a profession, forensic science must have peer-defined standards of competence, a code of ethics, a body of specialized knowledge and a strong educational core. Unfortunately, there are few forensic science programs, and very few that are considered to be grounded in a strong natural science background. Consequently, there are many nonforensic graduates working in crime laboratories. These individuals receive on-the-job training in order to acquire the necessary technical skills for their job and are often not able to acquire criminal justice philosophy and practice, and forensic science ethics, attitudes and perspective. Forensic science education can address these topics in its curriculum.

Along with the question of whether or not forensic science is a profession is the concern as to whether or not it is, in fact, a science. Practitioners, as expected, agree that it is both. Users of forensic science, and the general public, however, may not think so. Nonetheless, a forensic scientist's hallmark is an attempt to determine how events of legal significance occurred, based on observation, the formation of hypotheses and experimentation. This can easily be interpreted as a definition of science.

Unlike criminalistics, the other forensic sciences such as toxicology, pathology, anthropology and odontology have long-established education programs leading to an advanced doctoral degree. Consequently, this article is directed principally toward forensic science education as it applies to criminalistics (as defined and explained in detail in the article covering that field).

Education Surveys of Forensic Science Practitioners

Surveys of crime laboratory directors were carried out by Siegel in 1986 and Furton *et al.* in 1997. Others have written about the status of forensic science degree programs. Listings of such programs have been compiled by: (1) the Council of Forensic Science Educators (CFSE), an informal group organized under the auspices of the American Academy of Forensic Sciences (AAFS); (2) the United States Bureau of Alcohol, Tobacco and Firearms (BATF); and (3) the BC Institute of Technology (BCIT) in Canada. Some compilations of forensic science degree programs are available on the Internet.

Siegel's survey instrument was directed at three groups of respondents. One group, which comprised 23 forensic science majors at Michigan State University showed preference for: (1) a Bachelor of Science program in Forensic Science (BSFS) plus a Master of Science in Forensic Science (MSFS); or (2) a 5 year program leading to an MSFS. A second group consisted of 90 forensic scientists in the Michigan State Police, whose preference was an MSFS or a Master of Science in Physical or Biological Science. The third

group surveyed was made up of 153 members of the American Society of Crime Laboratory Directors (ASCLD). The preferred educational backgrounds for this group, although less focused than the other two groups, were similar. Siegel's survey led to a redirection of the Michigan State University forensic science offering, away from the baccalaureate degree and toward a graduate program at the Master's degree level.

It should be noted that surveys of forensic science practitioners, whether at the 'bench' level or at the management level, will necessarily be influenced by the educational background of the respondents and the workloads of most crime laboratories, especially in the United States. For many reasons, in particular the relative scarcity of forensic science programs, most laboratory directors came into the profession with a degree in a physical, more likely, or a biological science background. Added to this is the very heavy workload of most laboratories in the analysis of controlled substances, a forensic discipline that, more than any other, requires a strong chemistry background. Consequently, it is not surprising that the majority of practitioners and managers prefer an educational program that has a strong chemistry component. Furthermore, the current emphasis on DNA testing has caused the forensic science profession to search for analysts with a biology background.

Furton's survey asked 350 ASCLD members their preferred educational background, together with the specific courses they required of applicants. The respondents (52 out of 350 returned the survey) were told to answer the questions for each of the following disciplines, or combinations of disciplines: (1) drugs, (2) trace/impression evidence, (3) serology/DNA, and (4) firearms/documents/fingerprints. Each was also asked to identify his or her background. The authors concluded that the overall preference of the laboratory directors was for applicants who have a 'strong chemistry background'. Within each of the above-identified discipline areas, the preferred educational backgrounds were:

- drugs (63% chemistry, 15% biology, 13% forensic science, 9% other);
- trace/impressions (41% chemistry, 18% biology, 32% forensic science, 9% other);
- serology/DNA (28% chemistry, 41% biology, 21% forensic science, 10% other);
- firearms/documents/fingerprints (33% chemistry, 17% biology, 21% forensic science, 29% other).

The specialties of the laboratory directors responding to the survey were reported as:

- drugs/toxicology (40%);
- serology (13%);
- trace (21%);
- firearms/documents (16%);
- other (10%).

Again, the predominance of drug analysts as laboratory directors, and the likelihood that most of them had a 'strong chemistry background', would seem to skew the survey results.

Significance of an Education in Forensic Science

At the very least, forensic science is a scientific activity, i.e. it involves the use of the scientific method to describe, through physical evidence, how events occurred, based on observation, development of a hypothesis and subsequent experiments. Like medicine, forensic science responds to a specific case situation in each instance.

A science is based on relatively simple principles. The fundamental laws of chemistry, for example, are straightforward. So are the laws of biology, which deal with living organisms and vital processes. To arrive at a basic law of forensic science (again, referring primarily to criminalistics), one must understand the term 'identification' as it is used within the profession. It is the placing of an object within a class or group. There are many degrees of identification. As the class or group shrinks in size, 'identification' becomes 'individualization', and forensic science has its basic principle.

That 'individualization' is the defining principle of forensic science and, especially, criminalistics must not be dismissed as merely an obvious point or an inconsequential observation. On the contrary, the profession must realize that it is concerned only incidentally with identification as a prelude to individualization. The unfortunate confusion came about primarily from the development of methods for identifying people by their fingerprints, which resulted in so-called identification bureaus in many law enforcement agencies. Actually, persons were individualized by their fingerprints, not merely identified. Consequently, criminalistics began with a nomenclature that was inconsistent with science, and even today there is no universal acknowledgment of the field as a separate science whose basic principle is individuality.

Paul Kirk, in 1963, lamented the fact that 'the submergence of the concept of individuality by the very different concept of identification has retarded progress since it was not clear in which direction such progress lay'. He specifically contrasted the then lack of progress in the individualization of blood with the subdisciplines of firearms identification

and questioned document examination, neither of which stopped at simply identifying a make/model of gun or a group of possible writers. Rather, they went on to individualize the one and only gun, and the one and only writer. Kirk complained that serologists in the 1960s were content merely to identify the stain as human blood of a particular blood group. He went on to say that 'some (practitioners) may even go so far as to deny the possibility that blood is individual...' Were Paul Kirk alive today, it is certain that he would be among those loudly applauding the individualization of a bloodstain by the analysis of DNA as having originated from one, and only one, person.

But before we conclude that the individualization of blood, and other body fluids and parts, via DNA means that we have reached our goal, let us remember that glass, paint, soil, fibers, paper, ink and an infinite list of physical evidence types, including hair, are still out there in the identification phase.

Do chemistry or biology programs feature the science of individualization as a core of their course offerings? The answer is obvious. Only, at best indirectly, is it covered. Does this preclude graduates of such programs from becoming good forensic scientists? One need look no farther than Paul Kirk, who was educated as a biochemist, for an example. Nevertheless, if we are going to consider forensic science as a discipline, we must recognize that the principal entry route into a criminalistics laboratory is via specialized academic programs. The advantages are many. The forensic science major has: (1) selected a specific field and demonstrated the motivation and ability to handle the curriculum; (2) completed course work in subjects such as law, evidence, ethics and investigative approaches; (3) acquired basic scientific concepts with an emphasis on forensic science application; and (4) developed a forensic science ethic or way of thinking. The nonforensic science graduate, on the other hand, has to undergo a longer period of apprenticeship when employed. As effective as many on-the-job training programs are, they are not a substitute for an academic education program; any more than an apprenticeship can suffice for a doctor, dentist, pharmacist, or even a lawyer. Skill is not enough; for example, a skilled surgeon without an understanding of anatomy and physiology would be a menace.

A randomly selected survey of the baccalaureate degrees of members of the California Association of Criminalists (CAC) yielded the following results:

- forensic science/criminalistics (34%);
- biology/microbiology (34%);
- chemistry/biochemistry (24%);
- other (8%).

(Master's degrees were almost evenly divided between forensic science/criminalistics and others.) Because two-thirds of the practitioners in the survey apparently do not possess an educational background in forensic science, it is apparent that the responsibility for filling the education void has fallen on the profession itself, specifically on the training programs of criminalistics laboratories and on professional societies, especially certification and accreditation efforts that have evolved from them.

Notwithstanding the best training efforts of forensic science laboratories and the profession, university-based education has unquestioned benefits. Not the least of the benefits is the opportunity to carry out research, especially that which is discipline oriented. It allows students to present their ideas for challenge and gives them the opportunity for problem solving that can prepare them for the real world. It is too easy for practitioners to rationalize that if the laboratory trains its employees in the skills necessary to perform their assigned tasks, they will have instilled a way of thinking that distinguishes a forensic scientist. The expectation that the laboratory analyst will know how to approach a problem and ask the right questions, and use science wisely and ethically within the legal system, is unrealistic.

Content of Forensic Science Education

An advantage of a forensic science curriculum, perhaps more so at the graduate level, is that it can allow a diverse program with many courses that are potentially valuable to the practicing criminalist. Examples include criminal investigation, microscopy, physical anthropology, genetics, geology, law and research methods.

One group of authors recommended a program with a well-defined and relatively intense research component, which can be provided at the graduate level. Others suggested a formal internship program with a set schedule of attendance and activities, an appropriate length of time, specific assignment(s) and an evaluation of performance. Yet another urged instruction of ethical considerations in a forensic setting.

Two industry surveys revealed results that were not surprising. One, in 1976, showed the preferred course – offering blocks to be (in order of preference) criminalistics/microanalysis, instrumental analysis, serology/immunology and an internship. The second, in 1998, indicated an across-the-board (for each of the authors' identified subdisciplines) preference for chemistry, biology and mathematics/statistics.

Colleges and Universities that Offer Forensic Science Programs

College and university programs in forensic science are available internationally, although the majority are underfunded and often limited to only a few students. They provide educational opportunities principally at undergraduate and Master's degree levels. Outside the United States, forensic science education is more likely to be in a forensic medicine setting and to include a 'working' laboratory component that provides services to a regional criminal justice system.

Doctoral forensic science education in the United States is generally an interdisciplinary program. An example is the PhD offered by the City University of New York's John Jay College in Criminal Justice with a concentration in forensic science. Internationally, there exist at least two formal PhD programs. One is at the University of Strathclyde in Scotland; the other, at the University of Lausanne in Switzerland.

Listings of colleges and universities that offer forensic science programs are available on the Internet. The American Academy of Forensic Sciences (AAFS) web site has the following programs (with contact names, addresses, telephone numbers and other information shown for most of the institutions).

Undergraduate programs (United States)

Bachelor of Science in Forensic Science

Eastern Kentucky University, Richmond, Kentucky
George Washington University, Washington, DC
John Jay College of Criminal Justice, New York, New York
Michigan State University, East Lansing, Michigan
University of Alabama at Birmingham, Birmingham, Alabama
University of Central Florida, Orlando, Florida
University of Central Oklahoma, Edmund, Oklahoma
University of Mississippi, University, Mississippi
University of New Haven, West Haven, Connecticut

Bachelor of Science in Chemistry with Concentration in Criminalistics

Metropolitan State College, Denver, Colorado
York College of Pennsylvania, York, Pennsylvania

Bachelor of Science in Chemistry with Emphasis in Forensic Science

Ohio University, Athens, Ohio

Bachelor of Science in Forensic and Toxicological Chemistry

West Chester University, West Chester, Pennsylvania

Dual Major, Chemistry and Criminal Justice

Weber State University, Ogden, Utah

Forensic Science or Criminalistics Concentration in Criminal Justice Program

Albany State University, Albany, Georgia
Weber State University, Ogden, Utah

Bachelor of Science in Forensic Investigation

Jacksonville State University, Jacksonville, Alabama

Minor in Forensic Science (within Polymer Science Department)

University of Southern Mississippi, Hattiesburg, Mississippi

Minor in Forensic Science (within Criminal Justice)

St John's University, Jamaica, New York
York College of Pennsylvania, York, Pennsylvania

Undergraduate programs (outside the United States)

Australia: The University of Technology, Sydney, New South Wales
Canada: Laurentian University, Sudbury, Ontario
Germany: Humboldt University of Berlin, Berlin, Germany
Philippines: Philippine College of Criminology, Manila, Philippines
Poland: University of Krakow, Krakow, Poland
Switzerland: Universite de Lausanne, Lausanne, Switzerland
United Kingdom: University of Glasgow, Glasgow, Scotland
University of Strathclyde, Glasgow, Scotland

Graduate programs (United States)

Master of Science in Forensic Science

University of Alabama at Birmingham, Birmingham, Alabama
George Washington University, Washington, DC
John Jay College of Criminal Justice, New York, New York
University of New Haven, West Haven, Connecticut

Master of Science in Criminalistics

California State University, Los Angeles, Los Angeles, California
California State University, Sacramento, Sacramento, California
University of Illinois at Chicago, Chicago, Illinois

Master of Forensic Science/Master of Science in Toxicology
George Washington University, Washington, DC

Master of Science in Criminal Justice/Concentration in Forensic Science
Albany State University, Albany, Georgia
Michigan State University, East Lansing, Michigan
Virginia Commonwealth University, Richmond, Virginia

Master of Science in Chemistry – Forensic Related Research
West Chester University, West Chester, Pennsylvania

Master of Science in Natural Science – Forensic Concentration
Southeast Missouri State University, Cape Girardeau, Missouri

Graduate programs (outside the United States)

Master of Science in Forensic Science
India: Punjabi University, Patiala 147 002
University of Madras, Chepaulz, Chennai 600 005
University of Sagar, Sagar (M.P.)

Master of Science in Criminology
Philippines: Philippine College of Criminology, Manila, Philippines
Switzerland: University of Lausanne, Lausanne, Switzerland
United Kingdom: King's College London, London, England
University of Strathclyde, Glasgow, Scotland

Master of Science in Clinical Pathology and Clinical Forensic Medicine ... in Forensic Toxicology ... in Facial Identification ... in DNA Analysis ... in Human Identification
United Kingdom: University of Glasgow, Glasgow, Scotland

Doctoral programs (United States)

PhD in Criminal Justice with Concentration in Forensic Science
John Jay College of Criminal Justice, New York, New York

Doctor of Public Health with Research Opportunity in Forensic Science
University of California at Berkeley, Berkeley, California

PhD in Toxicology with Forensic Emphasis
University of Illinois at Chicago, Chicago, Illinois

Doctoral programs (outside the United States)

United Kingdom: University of Glasgow, Glasgow, Scotland
Switzerland: University of Lausanne, Lausanne, Switzerland

Undergraduate and graduate courses in forensic odontology

Louisiana State University, New Orleans, Louisiana
Northwestern University, Chicago, Illinois
University of Louisville School of Dentistry, Louisville, Kentucky

Courses recognized by the AAFS Odontology Section (as fulfilling the requirement of a formal course of instruction)

American Board of Forensic Odontology, Colorado Springs, Colorado
American Society of Forensic Odontology, Saratoga Springs, New York
Armed Forces Institute of Pathology, Washington, DC
University of Texas Health Science Center at San Antonio, Dental School, San Antonio, Texas

Formal Review or Accreditation of Forensic Science Education Programs

In the United States the American Chemical Society has the responsibility for reviewing and accrediting chemistry programs. It would seem reasonable that a professional organization representing forensic scientists could be charged with a similar task for its discipline. The AAFS is a logical candidate. It has almost 5000 members, representing eight forensic sciences. Approximately one-third of the membership is composed of crime laboratory practitioners. Furthermore, the organization has a significant international component, with some 360 members representing 58 countries outside the United States. The AAFS has a loosely knit group called the Council on Forensic Science Education that meets annually for a few hours as part of the Academy's annual meeting.

It is not an AAFS committee and, arguably because it is not officially sanctioned, has not taken any substantial steps toward some type of accreditation program. There exists the possibility that the Council may be able to work with certifying groups such as the American Board of Criminalistics (ABC), or accrediting bodies such as the American Society of Crime Laboratory Directors/Laboratory Accreditation Board (ASCLD/LAB), in an effort to conduct site visits to academic institutions that offer forensic science programs. At present there exist only lists, such as the previously cited one, that offer the prospective student a choice which is often based on geography, person knowledge or contact, word of mouth, or some other means of assessing the available forensic science programs.

Model for Forensic Science Education

Suggestions for improved forensic science education include: (1) the formation of regional cooperative centers; and (2) the placement of local, state/provincial and national forensic science laboratories on university campuses. The former would require mutual cooperation among programs within specific geographic areas and could be accomplished in several different ways. The latter is not novel and already exists in several parts of the world, although on an infrequent basis in many countries, such as the United States.

An interesting model has been proposed by one author, who examined the history of medical education during the period 1870 to 1926 and compared it to the current state of forensic science education. In 1870 medical education consisted of a 4 month term of lectures, followed by an *optional* period of unstructured apprenticeship and then professional practice. There were no entrance requirements, no written examinations, no laboratory exercises, no clinical work with patients and no university affiliations. The instructors were practicing physicians who taught during their free time. Medicine was not recognized as a science; there was no concept of medical experiment, testing or laboratory research. Soon, new physicians began pushing for something beyond the part-time lecture and apprenticeship system. At about this same time universities in the United States, like their counterparts in Europe, were developing as centers of excellence in fundamental learning. They soon began looking at medicine as an emerging science that needed an extensive, university-based education program. Harvard University, then Pennsylvania and Michigan, and to a greater degree Johns Hopkins, pioneered a change in medical education in the United States to essentially what it is today, i.e. an extensive multiyear program (following the baccalaureate degree) that features an association with a university-associated teaching hospital and clinical hands-on instruction.

Obviously, there are some obstacles to the pursuit of a similar model for forensic science. First, there are far fewer practitioners than in medicine. Second, crime *is* a major issue, but it would be difficult to compete with health care. Third, the relative importance of medical science in the context of health care exceeds the relative importance of forensic science to the justice system. Also, health care is not generally carried out within an adversary system. Nonetheless, this 'medical model' bears serious consideration for the many benefits it can bring to forensic science education; for example, an efficient combination of physical facilities, added personnel resources for both the laboratory and the university, access to research by the laboratory and help for the laboratory in its in-house training programs.

See also: **History**: Crime-scene Sciences. **Pathology**: Overview. **Nursing**.

Further Reading

Furton KG, Ya-Li Hsu and Cole MD (1999) What educational background do crime laboratory directors require from applicants. *Journal of Forensic Sciences* **44**:128–132.

Gaensslen RE and Lee HC (1987) Regional cooperation and regional centers among forensic science programs in the United States. *Journal of Forensic Sciences* **33**:1069–1070.

Kirk PL (1963) The ontogeny of criminalistics. *Journal of Criminal Law, Criminology and Police Science* **54**:235–238.

Lee HC and Gaensslen RE (1987) Forensic science laboratory/forensic science program cooperation and relationships: the view from the forensic laboratory. *Journal of Forensic Sciences* **33**:1071–1073.

Midkiff CR (1994) *International Education in Forensic Science (Part I. Graduate Programs, Part II. Undergraduate Programs)*. Scientific Sleuthing Review National Law Center: Washington.

Peterson JL (1987) Teaching ethics in a forensic science curriculum. *Journal of Forensic Sciences* **33**:1081–1085.

Peterson JL and DeForest DC (1976) The status of forensic science degree programs in the United States. *Journal of Forensic Sciences* **22**:17–33.

Sensbaugh GF (1997) On the advancement of forensic science and the role of the university. *Proceedings of the European Academy of Forensic Science*, Lausanne: Switzerland.

Siegel JA (1987) The appropriate educational background for entry level forensic scientists: a survey of practitioners. *Journal of Forensic Sciences* **33**:1065–1068.

Stoney DA (1987) A medical model for criminalistics education. *Journal of Forensic Sciences* **33**:1086–1094.

ELECTRONIC COMMUNICATION AND INFORMATION

M Y İşcan, Adli Tıp Enstitüsü, Istanbul Üniversitesi, Istanbul, Turkey
M Loots, Department of Anatomy, University of Pretoria, Pretoria, South Africa

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0766

Introduction

In the last 15 years or so, electronic communication has become one of the most widely spread and used aspects of modern technology. This technology has provided for rapid information exchange among many scientists. These exchanges include primarily: (1) correspondence between and among groups of individuals; (2) search for available information; and (3) exchange of scientific data. For example, a pathologist is looking for an opinion on the estimation of age from the ribs, the most reliable bone for this purpose. He poses the question on FORENS-L, a mailing list dealing with the forensic sciences which has nearly 1000 members. A knowledgable anthropologist responds by asking if he can see pictures of the rib in question. The pathologist scans several views of the sternal end of the rib and attaches them to an electronic message addressed to the anthropologist. Within a few hours the anthropologist e-mails back with an age estimate.

Electronic technology has changed so much that many interactions and tasks can be performed within the Internet with a click of a 'mouse'. The use of the Internet has become as common as the telephone or regular postal mail. Many scientific and popular journals, magazines, and newspapers are now available on-line to all those who have access to the Internet. There are even guides on how to conduct research on-line. The purpose of this article is to provide a brief understanding of what the Internet is and the resources available to and accessible by forensic scientists. It is important that forensic scientists are aware of these Internet resources to update and upgrade their professional skills, exchange ideas and new information, cooperate in research, and seek help from fellow scientists in resolving a particular medicolegal problem. This article provides a basic discussion on how this can be accomplished. There are numerous other publications for medical and scientific professions that provide a detailed introduction to many areas that are mentioned only briefly here. These include papers on various aspects of forensic sciences, biomedical fields, pathology, public health, biological anthropology and even bibliographic and reference resources on the Internet.

Internet

The Internet is a global system of computers connected via various networks and modems. Using standard protocols, these computers can communicate and actually exchange data. The Internet of today developed indirectly in response to the launching of Sputnik, the first artificial earth satellite, in 1957 by the then USSR. Because of this launch, the Advanced Research Projects Agency was started within the US Department of Defense in the late 1960s, and this led to the development of ARPANET. To enable computers on different networks and running different operating systems to communicate with each other, various protocols are used. The standard protocol used today (TCP/IP) was developed in 1973 and was adopted by all machines connected to ARPANET in 1983. Each computer on the Internet has a unique IP address. During this same year, the University of Wisconsin created the Domain Name System (DNS) to translate the name to an IP number. Instead of having to remember a cumbersome IP address which is composed of a series of numbers (e.g. 137.215.118.104), packets of data could now be sent to a domain name (e.g. maestro.up.ac.za) which is easier for humans to remember.

Internet service providers

There are various ways to gain access to the Internet. The most obvious is via a university network. Modem connections using an Internet service provider (ISP) are becoming faster and more affordable. When choosing a commercial ISP, it is important to look for a reliable local company that uses the latest technology and upgrades its system frequently. In some countries, the main cost is the monthly fixed fee, as in the US. In others such as South Africa and Turkey, there are usually additional charges (message units) from the phone company.

Internet software and repository sites

Software for all types of operating systems and personal computers can be found on the Internet. Although many are commercially available at a local computer shop, others can be obtained from remote computer sites. Some of these are freeware (no fee to use) and others are shareware. Shareware is distributed free of charge and can usually be evaluated for a limited time period after which it must be registered and a license fee is paid to the developers. Normal copyright restrictions still apply to both freeware and shareware. More and more commercial software is also being distributed via the Internet. Of these, many also supply a 'demo' copy for evaluation which expires after a certain time. Most of the free and shareware programs can be found in repository sites by anyone who has access to the Internet. Some of these sites are Garbo, Simtel, and Tucows. Others specialize in specific programs, like Windows 95/98 software or Apple products. There are also places where one can find software reviews. They can be reached via FTP or by using one of the many available web browsers.

Communication

Electronic communication forms one of the most important aspects of the Internet. As mentioned above, this communication provides for the exchange and distribution of messages, letters and documents (e-mail, listservers and usenet,) as well as real-time messages (on-line chatting) with another person.

Electronic mail and attachments

Electronic mail (e-mail) messages contain text sent from one person to another using computers connected via some form of network. This can be directly wired as in a local area network (LAN) or a dial-up system using a modem. E-mail is extremely fast, reliable and cheap. Most e-mail programs have a basic editor for composing messages. The message is sent to a recipient whose address must be specified. The message is stored in an electronic mailbox until the recipient retrieves the mail.

There are several popular mail programs, such as Netscape, Pegasus and Eudora, which can be found in many repository sites. Most mailing programs have various options such as CC, short for carbon copy, which allows the sender to send copies of the mail to other users. The recipient of the message should be able to see who the intended recipient is and who are getting the copies. In the case of BCC (blind carbon copy), a copy is again sent to each of the addresses but the recipient will not see the names of other recipients. This is especially useful when a message is sent to a large number of people as in a distribution list. Most mail programs also allow the user to add attachments to the message. Larger files (manuscripts, graphics, tables) can be sent as an encoded attachment to a normal e-mail message. These are encoded by the mailing program to enable it to pass through the various mailing systems.

If one is looking for a colleague's e-mail address, it can done in a number of ways. Zeno Geradts' forensic page is frequently updated and includes the most comprehensive list of forensic resources available on the Internet. Forensic scientists can add their details to this database. A more general searchable list of e-mail addresses is also available. However, the material in some of these general databases might not be very current.

There are many commercial companies as well as regional community networks that provide free e-mail accounts. In the US, for example, http://www.juno.com offers free e-mail accounts and the dial-up software. It has most of the basic functionality of normal e-mail. People with erratic or shared access to the Internet can still get a free e-mail address from a variety of places. These are all based on the web and are financed by advertising on their webpages. They can all be accessed from anywhere in the world as long as the user has access to the World Wide Web.

Listservers and mailing lists

A mailing list is a list of e-mail addresses identified by a single name. When a message is sent to the mailing list address, a listserver automatically forwards the message to all the addresses on the list. This automated system allows a user to send one message to the central address to be distributed to all subscribers on the list. People from all over the world can participate in discussion on a topic of mutual interest (e.g. Forens-L, a discussion list on forensic sciences).

A list usually has two addresses associated with it. The first is the address to which all commands regarding the subscription should be send to. This address is used to subscribe to the list, update or change subscription options or to sign off from the list. The second address is the address of the list itself. All messages sent to this address are remailed/distributed to all subscribers on the list. Most lists also have owners or moderators (e.g. M.Y. İşcan is founder and owner of Forens-L), and sometimes a technical manager. To subscribe to a mailing list, a 'subscribe' command is sent to the listserver. The command is placed in the body of the e-mail. The syntax for all systems is basically the same. If the syntax is wrong, it

is very likely that a help message may follow in the returned mail. The exact format can be deduced from the address of the server:

LISTSERV (or MAILSERV)
  subscribe listname yourfirstname yourlastname
LISTPROC
  subscribe listname yourfirstname yourlastname
MAILBASE
  join listname yourfirstname yourlastname
MAJORDOMO
  subscribe listname

A list of lists can be found at http://www.liszt.com containing more than 80 000 addresses of discussion groups. It provides a search function as well as an opportunity for list owners to submit their lists to the database. A list can be open for anyone to join, or closed, requiring permission to join. It can also be moderated. In this case the list owner/moderator approves messages of value and eliminates 'junk' mail before distribution to list members. Under the key word 'forensic' there are several lists including the following:

Clinical Forensic Nursing
listserv@ulkyvm.louisville.edu
  subscribe CLFORNSG
Council on Forensic Science Education:
listproc@lists.fsu.edu
  subscribe COFSE YOUR FULLNAME
Forensic Psychiatric Issues:
listserv@maelstrom.stjohns.edu
  subscribe FORENSIC-PSYCH
Professionals in the Forensic Sciences:
listserv@nervm.nerdc.ufl.edu
  subscribe FORENSI

Most e-mail programs also have a built-in mailing list function which makes it easy to send a message to selected colleagues and such a list can be formed in a few minutes. For example, the mailing list function of Pegasus Mail, combined with various filtering options, can be used to set up a basic but functional mailing list.

Usenet

Usenet is a worldwide bulletin board system. There are over 14 000 newsgroups that cover a wide range of interest groups used daily by millions of people worldwide. These newsgroups can be accessed through the Internet or an on-line service (without restrictions) to access or to post a message. Unlike mailing lists, Usenet messages are posted to specific newsgroups instead of being distributed to the members of a mailing list. Only the local system administrator determines what groups should be carried by their system.

Mainframe users can access the newsgroup with various newsreader programs including rn and tin. A variety of programs are also available (some integrated into web-browsers) for users of graphical operating systems. Some systems do not provide direct access to Usenet. In this case, newsgroups can be read at other sites. This free service enables registered users to access their favorite forums from any computer with a web browser and post messages via a web-based, spam-filtered e-mail service. A large selection of newsgroups are also archived here and can be searched by keywords.

Newsgroups are organized in a hierarchical system according to the area of interest. The following major categories exist:

1. comp: computer science, software and information on hardware and software systems;
2. misc: topics that do not fit into any of the other headings or cover multiple categories;
3. sci: research in or application of established sciences;
4. soc: social issues, socializing and discussion of different world cultures;
5. talk: debate-orientated with a high noise ratio;
6. rec: groups oriented towards hobbies and recreational activities.

There are also a number of 'alternative' hierarchies (alt, gnu and biz). In the alt-hierarchy a form of anarchy reigns and discussions vary from science fiction to evolution. Of the many available newsgroups, few have direct applicability for the forensic scientist. Sci.med.telemedicine and sci.psychology-hierarchy are useful groups to browse. There are, however, many other groups that can be a valuable source of information like the rec.photo- and comp.security-hierarchies or discussion on statistics (comp.soft-sys.stat.spss and comp.soft-sys.stat.systat).

Instant messaging

This is a form of real-time communication for two or more persons. Message services combine the speed of real-time conversation with the functionality of e-mail. Notes, web-addresses and files can instantly be sent to colleagues. Groups can chat without having to create large e-mail databases. Communication and collaboration can take place faster and cheaper than before. The one problem is still interoperatibility. The different services do not cross-communicate and the person who wants to 'talk' has to subscribe to the same service.

Two popular systems are ICQ by Mirabilis and Instant Messenger by AOL. By using this system, the user logs on to a central server. Friends and colleagues can be added to a list of known persons. When they are also on-line, the program alerts the users on both sides and a message can be sent directly to the recipient. Some systems allow users to search for people with similar interests and a chat-request can be sent to them. The whole session can be saved to a file for record keeping. The cost involved in this is mainly the local phone charges. International chatting is therefore very cost-effective for real-time communication.

Another form of chatting is known as Internet Relay Chat (IRC). This is probably the most widely used type of text-based real-time communication. Although the principle is the same as in personalized chatting, IRC is similar to a town hall meeting where anyone can join in the discussion. Anyone can set up their own IRC server and these servers can be connected to one another, giving the user access to all conversations around the globe. Almost all information needed for IRC can be found at http://irchelp.org, including primers, downloadable programs, technical guides and server lists. Liszt's also maintain a directory of IRC groups.

Remote Access

Another important aspect of the Internet is its ability to provide access to remote computer systems. Access is gained via Telnet and FTP. Although these functions are usually on a command line program, graphical interfaces are available for most computer systems making transfers to and from the remote system as easy as moving files on a personal computer. Speed of access does not depend on proximity. Many programs for these functions can be found in software repositories like Simtel and Garbo as mentioned above. There are many handbooks available that give a more detailed description of all these functions.

Telnet

Telnet is a program which is part of the TCP/IP protocol. Its purpose is to allow a user to log on to a computer from a remote location and execute commands. If allowed, one can also run any programs on the host machine using its processor and hardware with the local machine only acting as a terminal. This is most often used to read e-mail, access a library, weather information, and news services. Some of the mainframe programs like the SPSS statistical package can be accessed in this way. Assuming that the remote site supports Telnet, a simple sequence of Telnet commands is invoked by the initial 'telnet *hostname* (IP or site name)' line followed by user id and password.

File transfer protocol

File Transfer Protocol (FTP) is also part of the TCP/IP to allow the transfer of larger files between two machines connected via the Internet. Only basic commands to facilitate the transfer of files are allowed. Most systems allow 'anonymous' access to the remote machine. With anonymous access, there is no need to have an account on the remote machine. Files can be transferred by anyone via FTP. The anonymous FTP directory is isolated from the rest of the system and will generally not accept uploads from users.

When using MS-DOS, UNIX or VMS to connect without the benefit of the graphical system, FTP is a command-line program. Although the principle is the same when using a graphical system, it is still useful to know what is happening in the background. Files can also be retrieved seamlessly with a web browser. To connect to a site without the aid of a graphical system, use the command:

ftp *hostname* or, if the ftp-client is already running,

open *hostname*, for example

open *maestro.up.ac.za*

or open *137.215.118.104*

If there is a problem with the hostname, the numerical IP address can also be used. When a connection is established, a message from the remote system will first request a username and then a password. As soon as a connection is established, the remote system asks for

User (maestro.up.ac.za:(none)): *anonymous*

331 User name OK, please send complete E-mail address as password

Password: *guests* (or the user's e-mail address)

230 User ANONYMOUS logged in

ftp>

At the ftp-prompt, the user can now type *help* or *?* to receive a list of available commands. The most often used commands are:

ascii: Tell the machine that the file to be transferred is in ascii-format

binary: Files to be transfered are not in ascii format. This is used before transferring documents and programs.

dir: Display a list of files and directory on the remote machine.

get: Transfer files from the remote server to the local machine.

put: Transfer files to the remote machine

Mput and *mget*: Transfer multiple files in one command.

At the end of the session, a '*close*,' '*quit*' or '*end*' command is used to close the connection between the two machines. These commands are analogous to UNIX commands but the exact syntax might vary from system to system. An explanation of the technical side of FTP can be found at http://www.scit.wlv.ac.uk/~jphb/comms/ftp.html. There is also a detailed tutorial page at http://www.uic.edu/depts/adn/infwww/inform/vftp.html.

Information Retrieval

World Wide Web

Many of the Internet-related utilities such as Telnet and FTP can now be handled via the World Wide Web (WWW). Most of the browsers are supplemented with these programs. In 1990, CERN in Geneva implemented a hypertext system to provide more efficient access to information for its members of the international high-energy physics community. Two years later, the WWW was released by CERN. Since its inception, the WWW has grown beyond all expectations.

Along with this service came the problem of locating information on the web. Various 'search engines' have been developed. Studies comparing six major search engines and estimating the number of web sites indicated that the size of the indexable web is at least 320 million pages. Search engines make it possible to look for Internet resources on a specific subject. A list of available search engines can be found and also a more practical introduction to the capability of a search engine. Most of the search engines are specialized and may give different results for the same search words so it is advisable to try several systems. Yahoo is one of the best known search engines. There are also sites covering a specific region, e.g. the Dutch Electronic Subject Service (DutchESS) or academic information. A general multipurpose search engine is Dogpile which includes many of the most popular sites such as Yahoo!, Lycos' A2Z, Lycos WebCrawler, Excite Guide, GoTo.com, PlanetSearch, Thunderstone, What U Seek, Magellan, InfoSeek, Excite and AltaVista. On the same site, one can also search Reference and Dejanews for information available in Usenet.

Most forensic science sites can be found by using these search engines. Zeno's Forensic Page is one of the most comprehensive and up-to-date resources on forensic material available on the Internet. It includes links to numerous scientific organizations, journals and books, and lists upcoming conferences and meetings. There is also a comprehensive database of e-mail addresses of people involved in forensic sciences.

In addition, information about a number of forensic scientific journals is available online. There are also many journals available online and more are being added each day. A list of available electronic journals can be seen at http://www.edoc.com/ejournal.

Webdesign

The WWW has become increasingly pervasive in society, covering all aspects of everyday life including commerce, leisure and education. One of the key features of the web is the ease with which individuals can publish information, and the simplicity with which sophisticated and exciting results can be achieved in a short period of time. Almost any kind of subject can be found on the web, and it is often of very high quality.

Writing a personal page is not as difficult as it might seem at first and it is even easier to work with the basic hypertext markup language (HTML) codes than using a wordprocessor with a built-in HTML function. There are numerous HTML programs some of which are free public domain. One of the friendlier, free for nonprofit or personal use is Web-O-Rama. Garbo and Simtel are again good places to start looking for others. Many ISPs carry some of the popular programs or their own specific outputs.

Once created, a personal web page can be linked to other pages on the web to form a web site. There are several companies that provide free space to design a personal web page. Many of these also provide a web-based e-mail address but not Internet access. An overview of the most common pitfalls and errors in designing an effective page can be found.

Conclusion

Electronic communication has become a must for an active forensic scientist to search for new development in his field of specialization. The Internet, the platform where all electronic activities between separate workstations meet, is new and yet it has significantly facilitated progress in the daily life of the modern world. Many medical experiments and even surgery can now be performed in front of an Internet-connected video camera to share the experience of one set of physicians with others. Information available in all parts of the world from China to Cuba can be accessed

by anyone who has a computer and access to a network. Literature is not only available through Internet-connected libraries, but can also be received via fax (at low cost and in a short time) once located in a library that has this facility. Even scientific journals are providing FTP locations for contributors to upload new manuscripts to be considered for publication.

The main function of the Internet can be summarized in three different areas: E-mail and its variant forms like listservers and Usenet are the most commonly used offering. For a more sophisticated user and active researcher, FTP and Telnet are the most useful Internet functions. As was the case for the manuscript of this paper, data can be transferred exactly as is from one site to another, e.g. back and forth from South Africa to the USA. For information seekers, the WWW is a place that can easily be searched for new ideas, either by a direct visit to a site or through a search engine.

See also: **Clinical Forensic Medicine**: Sexual Assault and Semen Persistence.

Further Reading

Chamakura RP (1997) Internet and forensic science: current utilization and future potential. *Forensic Science Review* **9**:97–122.

Coiera E (1997) *Guide to Medical Informatics, the Internet and Telemedicine*. London: Chapman & Hall.

Dito WR (1998) Informatics. Using the Internet for forensic sciences. *Clinical and Laboratory Medicine* **18**:91–97.

Gralla P and Reimer M (1997) *How the Internet Works: Special Edition*. Emeryville, CA: Ziff Davis Press.

Hahn H (1996) *The Internet Complete Reference*. Berkeley, CA: Osborn McGraw-Hill.

İşcan MY and King WR (1997) Electronic communication for biological anthropologists. In: Boaz NT and Wolfe LD (eds) *Biological Anthropology: The State of the Science*, pp. 195–206. Corvalis, OH: Oregon State University Press.

Kastin S and Wexler J (1998) Bioinformatics: searching the net. *Seminars in Nuclear Medicine* **28**:177–187.

Lowe HJ, Lomax E and Polonkey SE (1996) The World Wide Web: A review of an emerging internet-based technology for the distribution of biomedical information. *Journal of the American Medical Information Association* **3**:1–14.

Lugo-Vicente HL (1997) Role of Internet in medicine. *Boletin-Asociacion Medica de Puerto Rico* **89**:82–87.

Pajdak W (1997) Why not benefit from Internet resources: a list of useful WWW addresses. *Acta Biochimica Polonica* **44**:363–364.

Pouliquen B, Riou C, Denier P, Fresnel A, Delamarre D and Le Beux P (1995) Using World Wide Web multimedia in medicine. *Medinfo* **8**:1519–1523.

Electrophoresis *see* **Analytical Techniques**: Capillary Electrophoresis in Forensic Biology; Capillary Electrophoresis in Forensic Science.

ENGINEERING

D Rudram, Forensic Science Service, London, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0541

Introduction

Forensic engineering is the application of engineering knowledge and principles in a legal context. While many practitioners will be involved in the investigation of engineering failures and the subsequent litigation, this article will discuss their contribution to other types of legal investigation. Forensic engineering covers the full gamut of engineering practise, from the investigation of a road traffic incident by a mechanical engineer to the recovery of data from a computer system by an electronic engineer. It is as broad a topic in its own right as forensic science.

The boundary between science and engineering is diffuse, although most people would have no difficulty in distinguishing between an engineer who designs car engines and the scientist who investigates

flame propagation in compressed gases, despite the fact that their work might well be linked. The dictionary is little help; it defines engineering as:

> The application of science for directly useful purposes.

The distinction becomes even more blurred when we try to differentiate between that most applied of sciences, forensic science, and forensic engineering. Is the practitioner investigating the behavior of skidding vehicles a physicist or a mechanical engineer? In reality we are discussing labels rather than content, and the crucial similarity is the application of technical and scientific knowledge to legal problems.

This article takes a broad look at how engineering disciplines are able to help the legal process, both criminal and civil. A number of forensic engineering topics have been dealt with in depth elsewhere in this encyclopedia and those articles provide an excellent way to start deepening your knowledge of particular areas.

Accident Investigation

The specialist role of the forensic engineer is most obvious in the investigation of accidents. These may be accidents, if that is the right word, involving any mode of transport; the workplace or machinery. The word 'accident' is unfortunate, as it suggests that the incident was unavoidable and no one's fault; in many, if not all, cases the investigation will show that the incident was both avoidable and the result of the actions of one or more of the parties. In any event the investigator will have to combine the results of his or her own observations, measurements and tests with information from witnesses, automatic recording equipment, television monitors and the known behavior and performance of the vehicle, equipment or materials involved. For example, following a road traffic accident the forensic engineers may have information from the scene, witnesses, a tachograph (a form of recording speedometer) and traffic management cameras, plus the results of their own observations and measurements to consider. They must combine these into a coherent whole that gives proper weight to all the evidence and which ignores or distorts none of it.

The use of the word 'accident' to cover crashes and other transport disasters is interesting in itself. It normally implies an unexpected or unintended event. A large number of transport accidents, particularly on the roads, are the entirely predictable outcome of deliberately chosen behavior. There is nothing accidental about them at all and the trend towards referring to them as incidents or crashes is welcome. The phrase 'dangerous occurrence', which comes from health and safety regulations, is particularly apposite.

Engineering can help to provide answers in the investigation of these incidents in a number of ways; three of these are discussed in this section.

Mechanical engineering

A critical component in the investigation of an accident or other dangerous occurrence is the behavior of the systems involved. As a result, much time is spent in collecting and examining wreckage. Often great efforts will be made to recover wreckage from locations as inaccessible as the seabed (**Fig. 1**), or to collect and reassemble small pieces spread over a large area, as happens when an aircraft breaks up in the air. The interaction between a vehicle and terrain is another major source of information. As with all other scene examination, detailed observations need to be made and recorded. The initial investigation must be conducted with no preconceptions about what is important and what is not.

Examination of the wreckage can tell the forensic engineer the attitude of the vehicle as it crashed. For example, did the car strike the wall frontwards or backwards; or was the nose of the aircraft pitched up or down as it struck an obstacle? Other witness marks may indicate that parts of an aircraft broke off before the final impact, enabling the investigator to begin to establish a sequence of events. The nature and location of damage, coupled with contact traces such as smears of paint, will enable the investigator to deduce how two cars may have collided. The nature and extent of the damage to the vehicle is a function of the severity of the impact and it is possible to work backwards from the damage to an estimate of impact speed, or at least speed change on impact.

Marks left as the vehicle interacts with the terrain can provide a whole range of information (**Fig. 2**):

- Skid and scuff marks left by a car as it goes out of control can not only provide information about the movement of the vehicle but also enable its speed to be estimated.
- The gouges left by an aeroplane propeller can indicate that the engine was still under power as the plane struck the ground.
- Marks caused by the flanges of a wheel on the side or head of a rail will show where a train started to derail.

Successful investigation requires the forensic engineer to combine all sources of information to produce a comprehensive reconstruction of events. Thus, while impact damage alone does not allow the investigator to calculate the impact speed of a car, couple this with

Figure 1 (see color plate 19) The fishing vessel Pescado was recovered from the sea off the south west coast of England to assist the Marine Accident Investigation Branch enquiry into the loss of the vessel and crew. Marine Accident Report 1/98 (1998), London DETR. (Copyright – The Chief Constable, Devon and Cornwall Constabulary.)

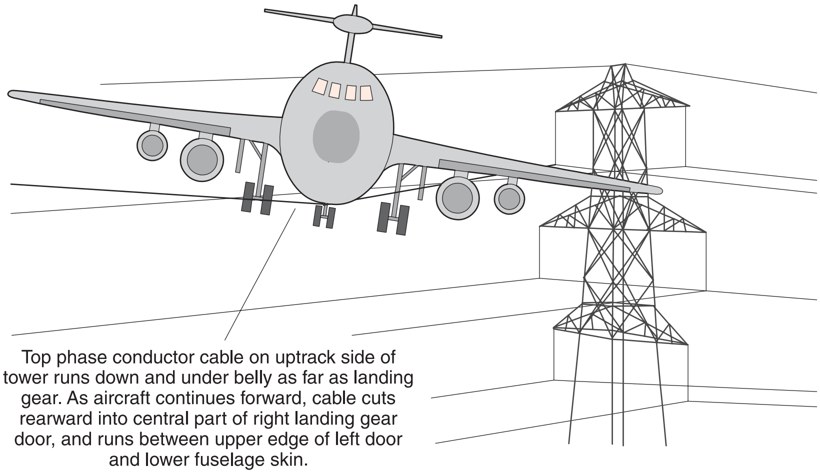


Figure 2 Witness marks on the underside of a Boeing 737 enabled investigators to deduce its attitude as it struck overhead electricity cables near Coventry, England on 21 December 1994. Air Accident Report 1/96 (1996). London: HMSO. Crown Copyright.

evidence about the post-impact movements of the vehicle obtained from tire marks and reliable information about impact speed may be obtained.

Defects and failures

Catastrophic failure of systems and components can precipitate an accident or other dangerous occurrence. Allegations of failure may be suggested as an explanation for an incident and it is critical that the forensic engineer is able to distinguish between post-accident failure and accident-induced damage. Road traffic incidents are a good example of where one or other party has claimed that a component failure was the cause – for example, a tire blowout – where careful examination has shown that the failure was a consequence and not a cause of the incident.

The type of failure that can precipitate an incident will depend on the type of transport. Thus, sudden engine failure is rarely a problem for a road vehicle but can of itself cause an aircraft to crash. As a result, considerable effort will be made to recover the engines and associated components from a crashed aircraft and subject them to meticulous examination. Fatigue failure of components is still causing aircraft accidents, and suggestion that fatigue may have played a part in an incident will result in a detailed review of the history of that component.

The system of air worthiness certificates and registration makes unauthorized modifications to aircraft unusual, if not impossible. Unfortunately, they contribute to a number of road traffic incidents every year. Modifications to the springs and dampers which cause unanticipated changes to the ride and handling of the car is one common form that can prevent the driver retaining control of the vehicle. Investigation of this kind of incident should extend beyond examination of the altered components to a practical investigation of their effect on the handling of the vehicle. Such tests must be conducted off the public road in an area where there is plenty of space for the driver to regain control of the vehicle.

Signaling and control systems

In an increasingly crowded world all modes of transport are controlled in some way. Controls can be as simple as a set of rules, such as are contained in national highway or driving codes. In controlled airspace, however, all movements of aircraft are monitored and directed by the air traffic controllers. Aircraft movements are controlled by a complex mix of verbal instructions and electronic surveillance. Signals are an alternative way of controlling movements; they may be as simple as a fixed STOP sign that instructs a motorist to stop at a junction, or as complex as the systems on some railways, which can detect and control the movements of trains.

Without reliable signaling and control the volume of traffic that can pass safely through the busiest parts of a transport network would be severely curtailed. Vehicle operators come to rely on the information the signaling and control system provides. Failure of the system in any way can have catastrophic effects. After any serious incident, part of the subsequent inquiry will be directed to the functioning of the signaling and control system. This can be a complex task, particularly when the system can alter and adapt to traffic conditions. For example, the automatic traffic signals controlling a complex road junction will have phases that are only activated when a particular set of circumstances apply. The engineer investigating the system needs to appreciate this and have sufficient reliable information to be able to trace the particular set of circumstances that obtained at the time of the incident.

Like many others, the weak link in signaling and control systems is all too often the humans operating it. Human failures have compromised the safety features built into railway signaling systems many times over the years, and even the highly structured environment of air traffic control does not eliminate human failing completely. In fact, much of the complexity of signaling and control systems is an attempt to eliminate scope for human failing.

Animation and simulation

Simulation is the use of a theoretical model of a system to reconstruct (or predict) events. Simulation techniques have been used in the design and development of all forms of transport but it is only with the advent of powerful personal computers that it has been a practical proposition to use simulation for the investigation of road traffic incidents. Of course, the very sophisticated flight simulators used for pilot training have been used to explore the behavior of an aircraft in an unusual flight regime as an aid to aircraft accident investigation for many years.

While simulation programs can produce purely numerical data, their great advantage is that this data can be used to generate an animated reconstruction of events. The use of animated output makes the result of the simulation much more accessible to nonspecialists and can be used to demonstrate the reconstructed events to a judge or jury. Animation is such a powerful tool for explaining a theory that the engineer using it must be careful to ensure that the medium does not become the message. Often a simple wire frame or polygonal model will suffice and has the advantage that the audience will never lose sight of the fact that

they are watching a reconstruction. The other important fact that must never be overlooked is that the critical component of any animation is the underlying simulation model that has been used to generate it. Characters in animated cartoons regularly do physically impossible things; here the underlying model is not a reliable representation of reality.

Practical modeling systems are a compromise brought about by the limitations of the available computing power and the amount of data that can be collected. Thus CRASH III, which in one of several personal computer implementations is among the most widely used programs for modeling the behavior of colliding cars ignores inter-vehicle friction completely. This program assumes that the contact points on the two vehicles achieve a common velocity; it is quite unable to handle sideswipe impacts.

Full simulation of the behavior of motor vehicles requires a three-dimensional model with 16 degrees of freedom. This requires a great deal of computing power, and frequently more knowledge of events than can be obtained. An all encompassing simulation would require information about tire cornering characteristics, tire–road friction, steer angle, load, driving/braking forces, suspension and damping properties for each wheel – over 30 different parameters. PCCRASH, a somewhat simpler model, can be used to good effect in many cases. This will allow for lateral, longitudinal and yaw displacement and can allow for the effect of pitch and roll on dynamic weight distribution. Good results can be obtained using a simplified data set; thus, steer angle and tire–road friction can be entered for each wheel but damping forces and tire cornering characteristics can be ignored.

One of the most challenging simulation problems in transport accidents is modeling the behavior of the human body. Complex models of the human body have been used to reconstruct the behavior of victims in a crash. Different human simulation models were used to investigate the behavior of victims of a crash involving a Boeing 737 in central England in 1989 and a railway accident in London in 1991. The analysis of the latter, which involved a rush-hour train at a terminal station, also had some interesting human factors. Standing passengers were more likely to be seriously injured than those still seated, and passengers in the rush hour start moving towards the doors as the train pulls into the station, leaving them more vulnerable if it then collides with the buffers.

Ergonomics

Whether they are called accidents, incidents or dangerous occurrences, any investigator soon becomes aware of the importance of the human beings in the system. Around 90% of all road traffic incidents are the result of human failing in some way. Recognition that it was not enough to assign the cause of an incident to human error was first apparent in air accident investigation. One of the roots from which ergonomics eventually grew was research into the reasons for pilot error and what could be done to prevent it. However, concern with the effects of human failure and how it could be prevented was being expressed by the railway accident investigators in the nineteenth century. Simple reminders to signalmen not to change the signals, such as the mandatory presence of a member of a delayed locomotive's crew in the cabin, were introduced before track circuits and signal interlocking prevented it happening.

This concern for why mistakes are made is still a common feature of enquiries into rail and air accidents. When a Royal Air Force war plane and a helicopter collided in clear weather, the accident report included a discussion of why 'see and avoid' was not necessarily a sound basis for collision prevention in conflicts with fast jet aircraft. One, all too common, cause of rail accidents in the United Kingdom is a signal passed at danger. The interesting question, of course, is why they happen:

- Are there problems with brakes on the train or do modern disk brakes require different techniques than the older clasp brakes?
- Are the signals faulty in some way or do the drivers fail to see that they are at danger until too late?

In the latter case, expectation plays a part in the driver's failure to react. Frequently he or she will have been traveling under amber warning signals for some time and fails to register that the latest signal is now showing red for danger. The Railway Inspectorate, who investigate rail accidents in the United Kingdom, have returned to this subject several times and made a series of recommendations to minimize the impact of human factors on rail accidents.

The human component in a safety-critical engineering system is often the weak link. While human failure is an explanation for an incident, it is not an adequate diagnosis. If the event is not to be repeated, the forensic engineer needs to understand why the failure occurred and what can be done to prevent a repetition. If humans do not learn from experience they are destined to repeat their mistakes.

Electronic Engineering

The all pervasive nature of information technology and computers means that electronic engineers will become involved in all types of investigation. The

hardware on which modern signaling and control systems rely will be based on microprocessors and stored programs, and the skills of the software or electronic engineer will be as important as those of the mechanical engineer when investigating an incident involving such systems. Electronic and computer engineers have a particular contribution to make to the investigation when computers are involved or where electronic images are a potential source of evidence.

Computers

The advances in computer systems and telecommunications technologies have created increased opportunities for criminals to conduct their business. Such activities are often referred to as 'computer crimes' and encompass the three ways in which computers can be used for criminal purposes:

- When the computer may be the target of a crime. The criminal's aim is to cause damage to a computer system and its data.
- When the computer may be an 'instrument' used during an offence. For example, a computer can be used to carry out traditional crimes, such as fraud or theft, or may be used to produce illegal copies of credit cards while others may be used in the illegal cloning of mobile phones.
- When the computer is incidental to some other offence but is significant because it may be found during a scene examination and contain evidence of that, or other, criminal activity, and therefore has an 'intelligence' value.

The popular image of the computer criminal is of someone who breaks or 'hacks' into a computer system remotely and alters it to his or her advantage. In practice, however, a great many of these crimes are committed by people with legitimate access to the system.

Every examination of computers will be different because the computer media, its structure, the purpose of the examination, the elements of the crime and the type of information are different for each investigation. However, successful investigations require a good understanding of 'target' systems and the security measures in place, such as:

- How access to the system(s) is controlled:
  - whether physically by use of barriers or locked doors, etc.
  - logically by the use of passwords and user IDs
  - a combination of these.
- What sort of audit trails are maintained.
- If different levels of access are established:
  - what rights are associated with these
  - who controls access
  - who might be able to change access levels.

In theory, with this type of inquiry, access to the computer system and associated records will be provided by the owner or system manager. In practice, establishing who had access to the system at any particularly important time, what their access rights were and how these may have been exceeded will require a great deal of careful research and effort. Still one of the most common, and simplest, ways of gaining unauthorized access to computer systems is by obtaining the user ID and password of a colleague. This does not make the investigation process any easier.

While many aspects of these computer crime investigations will revolve around determining access levels and opportunities available to individuals, an entirely different set of skills is required in order to conduct a forensic examination of the computer systems themselves. Examiners will need to be prepared to examine anything from a personal organizer to a highly sophisticated and networked personal computer. In addition to being able to handle this range of hardware, they will also need to understand the particular operating systems in use on these systems and a variety of software packages that can be used to store and manipulate data. Knowledge of security aspects built into these products and how these can be circumvented is also essential to allow a comprehensive examination to be carried out.

Data held on a computer system are no different to information or text contained on a document. For this reason evidence that is to be based on a computer or on computer media is subject to the same rules and laws that apply to the documentary evidence. The doctrine of documentary evidence may be explained thus: 'The onus is on the prosecution to show to the Court that the evidence produced is no more or less than when it was first taken into the possession of police.'

When faced with a computer system that may contain vital information, the investigator should think carefully before doing anything that might affect the integrity of the data on the system and compromise the subsequent forensic examination process. Operating systems and other software frequently alter and add to the contents of the computer systems storage space. This is an automatic process without the user necessarily being aware that the data have been changed. Ideally, therefore, a copy should be made of the entire target computer system on to media that can be retained for examination and subsequent use in court. A number of recognized copying devices are available for this purpose. As always, a

full record of all work carried out during the examination must be maintained.

The protection of individual files on computer systems can be achieved in a number of ways. At its simplest, the owner of the file relies on the overall security of the system; however, once access has been obtained the file can be easily read. At the other extreme, a wide range of sophisticated encryption software is available which can be used to protect the information. In certain circumstances some routines can monitor the way in which the computer system is started and, if it detects any differences, will disable the system or even delete its contents. It should be noted that deletion of files can, in some circumstances, be reversed; however, recovering lost data is more difficult than protecting or preserving it in the first place.

Digital imaging

Digital imaging and the processing of digital images can be used to address a number of problems. As digital images do not require development or printing they are available immediately. At its simplest, this means that the photographer can modify the conditions to produce a satisfactory image before leaving the location of the incident. Moreover, a digital image of a fingermark or shoeprint can be sent electronically from the scene to the laboratory for comparison and identification before the inquiry team leave the scene of the incident.

One of the great advantages of a digital image is the ease with which it can be manipulated to improve the clarity. While this might involve something like contrast enhancement, which can also be done photographically, techniques such as deblurring, sharpening or background subtraction require manipulation of a digital image. The ease with which a digital image may be modified brings its own dangers. It is vital that the original image is retained unaltered; all changes should be made to a copy and a complete audit trail of everything done maintained. Parliament in the United Kingdom was sufficiently concerned about the scope for unscrupulous manipulation of digital images that a select committee was formed to address the whole issue of digital images as evidence.

Video cameras are widely used to monitor public areas for both crime prevention and traffic control purposes. Inevitably this means that criminal activity will be recorded on video tape. The quality of these images and the viewpoint from which they are taken make identification of suspects or what they are carrying difficult. Careful enhancement of the image may allow a person or his or her property to be identified.

The effects of perspective and parallax can give a false impression of the height or build of someone caught on these cameras. Geometric methods can be applied to the digital image to extract length and height information without a detailed knowledge of the optical characteristics of the camera used. Simple rectification programs can be used to prepare complete plans of a scene from a digital image, although, for these to work, at least one measurement in the photographed scene is required. If several images are available a three-dimensional representation of the scene can be produced. By using these techniques the movement of an object through the camera's field of view can be followed. If the time lapse between successive images is known, speed can be determined, which has obvious applications when investigating road traffic incidents.

The widespread use of surveillance cameras and the increasing power and sophistication of digital imaging techniques means that a great deal of evidential material will become available to the criminal justice system. Solving the problems associated with facial recognition will become essential if this mass of material is to be evaluated effectively. As has already been demonstrated with speed enforcement cameras, the resulting information has to be handled with sensitivity if the 'Big Brother' image of video surveillance is not to swamp its power as a crime prevention, reduction and investigation tool.

Fire Engineering

The investigation of fires is dealt with in detail elsewhere but it as an area of forensic engineering where an interdisplinary approach will pay dividends. Understanding the events that have given to rise to a fire, from an examination of the burnt-out remains of a building or vehicle, may involve chemists, physicists and engineers. Careful investigation of the fire scene is an essential part of the process, and the successful fire investigator will combine the practical investigative and problem-solving skills of an engineer with the ability to synthesize information from a wide variety of sources.

The basic process in fire investigation is to establish the point at which the fire originated and then deduce from the available evidence how the fire started. Before this process can begin, the investigator needs to gain safe access to the scene. Since the fire can disrupt both the structural stability of a building and its electrical supply, and produce noxious combustion products, a pause for reflection to decide upon a safe approach before commencing work should be

standard procedure. The process of investigation to determine the origin and cause of the fire can then start and be brought safely to a successful conclusion.

Where the development and spread of the fire needs to be established, computer simulation and model tests can be used. Both of these approaches were used to assist in the investigation of a very serious fire at King's Cross underground station, London, where they: 'first drew attention to an important and unsuspected phenomenon in the form of the trench effect.' (**Fig. 3**).

Conclusion

Forensic engineering has applications over a wide and increasing range of criminal and civil investigations; for example, the pervasive spread of computers and surveillance cameras means that electronic, computer and software engineering will become an important part of the investigation of many serious crimes. The skills and techniques that are used to investigate transport accidents can be of value in a range of investigations, not only those where a motor vehicle has been used as a weapon in an assault. Falls, nonaccidental injuries and stabbing are all areas in which forensic engineers have contributed to an investigation.

See also: **Accident Investigation**: Airbag Related Injuries and Deaths; Driver Versus Passenger in Motor Vehicle Collisions; Motor Vehicle; Rail; Determination of Cause: Reconstruction; Determination of Cause: Overview; Tachographs. **Computer Crime**. **Fire Investigation**: Types of Fire; Physics/Thermodynamics; Chemistry of Fire; Evidence Recovery at the Fire-scene; Fire-scene; Fire-scene Patterns; Laboratory. **Health and Safety**: Including Risk Assessment. **Photography and Digital Imaging**: Overview; Digital Imaging Enhancement, **Fingerprints (Dactyloscopy)**: Chemistry of Print Residue. **Internet Crime**.

Figure 3 Fire tests on scale models of the escalator troughs at Kings Cross were conducted by the Health and Safety Executive at Buxton to demonstrate how the flames lay low in the trough with a jet of flame at the top – the 'trench effect'. Investigation into the Kings Cross Underground Fire, Desmond Fennell QC Cmm 499 HMSO London (1988). (Crown Copyright.)

Further Reading

Bramble SK and Jackson GR (1994) Operational experience on image enhancement by frequency domain filtering. *Journal of Forensic Sciences* **39**: 920–932.

Brown JF and Obenski KS (1990) *Forensic Engineering – Reconstruction of Accidents*. Springfield, IL: CC Thomas.

Carper KL (ed.) (1989) *Forensic Engineering*. New York, NY: Elsevier.

Criminisi A, Zisserman A, Van Gool L, Bramble S and Comptone D (1998) A new approach to obtain height measurements from video (3576-40). In: Higgins K (ed.) *Investigation and Forensic Science Technologies*, Proceedings of Society of Photo-optical Instrumentation Engineers 3576.

Drysdale D (1985) *An Introduction to Fire Dynamics*. Chichester, UK: Wiley.

Fennell D (1988) *Investigation into the King's Cross Underground Fire*. Cmnd. 499. London, UK: HMSO.

Lux WJ (1995) *An Engineer in the Courtroom*. Pub. no. R-155. Warrendale, PA: Society of Automotive Engineers.

Noon RK (1992) *Introduction to Forensic Engineering*. Boca Raton, FL: CRC Press.

Noon RK (1994) *Engineering Analysis of Vehicular Accidents*. Boca Raton, FL: CRC Press.

Noon RK (1995) *Engineering Analysis of Fires and Explosions*. Boca Raton, FL: CRC Press.

Watts AJ, Atkinson DR and Hennessy CJ (1996) *Low Speed Automobile Accidents*. Tucson, AZ: Lawyer's & Judge's Publishing.

Wilding E (1997) *Computer Evidence*. London, UK: Sweet and Maxwell.

Enhancement *see* **Fingerprints (Dactyloscopy)**: Sequential Treatment and Enhancement.

ENTOMOLOGY

R W Merritt and **M J Higgins**, Department of Entomology, Michigan State University, East Lansing, MI, USA
J R Wallace, Department of Biology, Millersville University, Millersville, PA, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0499

Introduction

Entomology, the study of insects and related arthropods, is a relatively recent addition to the forensic investigator's toolkit. Because insects have predictable life histories, habitats and developmental stages, the presence or absence of certain species at a crime scene – particularly a homicide – can provide important clues regarding when, where and even how that crime occurred. The 1990s saw a resurgence of interest in forensic investigations by entomologists.

Because insects play an important role in the decomposition process, the category 'medicolegal forensic entomology' has been used to describe the arthropod involvement in events surrounding felonies, usually violent crimes such as murder, suicide and rape. While a corpse covered with fly larvae (maggots) may be offensive to most people, it should be remembered that organisms such as maggots and beetles fill an important ecological niche in removing dead animals from the landscape. That certain species of insects have evolved to specialize in feeding on dead and decaying flesh, and do so in a predictable manner, allows a forensic entomologist to determine how long a corpse has been exposed to insects. Most frequently, the forensic entomologist is asked to determine the interval between death and corpse discovery, referred to as the postmortem interval (PMI). Determination of the postmortem interval is based on two major principles: (1) that the colonization of a corpse by insects follows a reasonably predictable succession of species; and (2) that developmental time for each stage of each species of insect is also predictable. Both of these principles are heavily influenced by environmental conditions, particularly temperature.

Entomological Succession

When an animal dies, it goes through five successional stages of decomposition that coincide with the activity of bacteria and insects (**Table 1**). The duration

Table 1 Stages of decomposition, duration and associated arthropod fauna

Stage of decomposition	*Accumulated time*[a] *(days)*	*Common insects found*
Fresh		
Commences at the moment of death and ends when bloating is evident. The breakdown of proteins and carbohydrate molecules occurs during this stage	0–3	Adult blowflies, flesh flies, muscid flies, yellow jackets, ants, daddy longlegs
Bloated		
Putrefaction occurs during this stage. Gases produced from metabolic activities of the anaerobic bacteria cause the corpse to inflate	4–7	Adult and larval blowflies, flesh flies, muscid flies, rove beetles, hister beetles, carrion beetles, ants, assassin bugs, yellow jackets
Decay		
Begins when the abdominal wall is broken, allowing gases to escape and the corpse to deflate	8–18	Adult and larval blowflies, rove and hister beetles, carrion beetles, dermestid beetles, ants, scarab beetles, muscid flies, cockroaches, red-legged ham beetles
Postdecay		
In dry habitats, remains consist of dry skin, cartilage and bones. In wet habitats, a wet viscous material, termed byproducts of decomposition, are found around corpse	19–30	Dermestid and hister beetles, mites, fungus beetles, springtails, fungus gnats, fruit flies, cheese skipper flies, phorid flies
Dry		
Bones, hair, dry skin remain. Odor that of normal soil and litter	31+	Dermestid beetles, ants, cheese skipper flies, sow bugs, soldier flies, ground beetles

[a] Accumulated time estimated from studies on pig carcasses in southern Michigan, USA during late summer to early autumn.

of each stage is directly related to the air temperature to which the corpse is exposed, as will be discussed below. While the length of each stage varies with temperature, the series of stages itself is very predictable, and a forensic entomologist can estimate time of death based on the stage of decomposition, the insects present, and past temperature data. Adult flies are generally the first insects to arrive at a corpse, appearing within minutes or hours after death to lay eggs or live larvae in some instances. Several thousand eggs may be laid on a corpse in a short period of time and the resulting maggots are the most important group of insects on a corpse, for they will eventually consume most of the decaying flesh. Maggot activity, and thus decomposition of the corpse, is greatest through the decay stage (Table 1). Through this level of decomposition, an entomologist can determine the PMI with a high degree of confidence by looking at the developmental stages of the maggots present. After the decay stage, when most of the flesh is gone, changes in the corpse and the insect fauna occur at a much slower rate, and it becomes increasingly difficult for a forensic entomologist to determine the PMI accurately. These later successional stages are dominated by dermestid beetles (**Fig. 1**), carrion beetles (**Fig. 2**), and other insects that feed on the dried flesh, skin and hair that remain on the corpse.

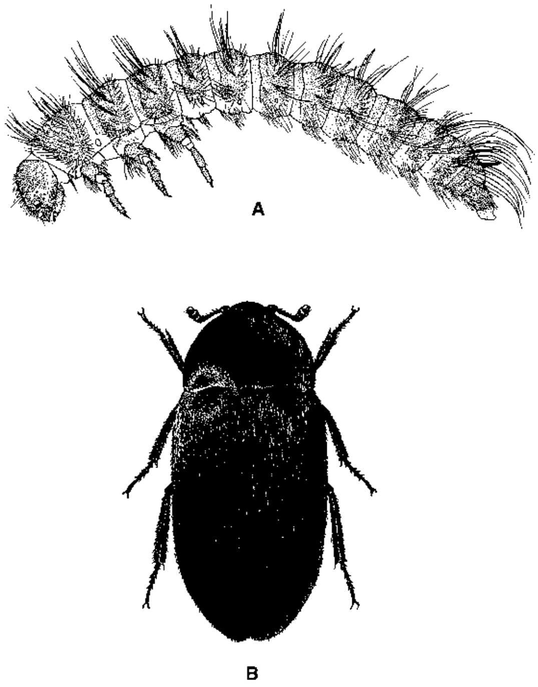


Figure 1 Dermestid beetles. **(A)** Larva; **(B)** adult. Reproduced from Gorman JR (ed.) (1987) *Insect and Mite Pests in Food: An Illustrated Key*. US Department of Agriculture, Handbook 655.

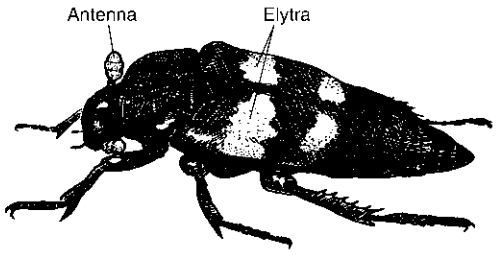


Figure 2 Adult carrion beetle, *Nicrophorus* sp. (20–25 mm in length). Reproduced with permission from Catts and Haskell (1990).

Life Cycle of a Fly

Without a doubt, the flies (order Diptera) are the most important group of insects that forensic entomologists use in determining the PMI. Two families of flies in particular have evolved to specialize on carrion: blowflies (Calliphoridae), and the flesh flies (Sarcophagidae). Blowflies are generally medium- to large-sized insects, with a metallic green or blue color, and are often referred to as 'greenbottle' and 'bluebottle' flies (**Fig. 3**). Adult female blow flies may begin to lay eggs on a corpse within minutes to hours after an individual dies. Egg laying, or oviposition, for these flies is primarily a daylight activity, and as such, may be delayed for several hours if a corpse is deposited during the night time. Eggs are initially deposited around naturally moist open areas of the body, such as eyes, nose, ears and mouth, but also are laid around wounds. A single adult female blowfly may lay up to several hundred eggs in a short period of time. Flesh flies are usually gray with black longitudinal stripes (**Fig. 4**). Female flesh flies deposit tiny live larvae in the same areas of the body where blowfly eggs are laid. The maggots of flesh flies are usually much less numerous than blow fly maggots.

The order Diptera undergo holometabolous metamorphosis (complete change) and with this type of metamorphosis the larvae or immatures are totally unlike the adults. Whether deposited as eggs or as small maggots, the larvae of flies go through three developmental forms, or instars, increasing dramatically in size during each instar before molting into the next instar (**Fig. 5**). The soft-bodied maggots feed on decaying flesh with hook-like mouthparts, and secrete enzymes that help them digest this food. The maggots breathe through slit-like structures at the posterior end of their bodies (posterior spiracles), and entomologists use the variations in these breathing structures and the shape of the internal larval head called the cephalopharyngeal skeleton to help them identify species (**Fig. 6**). As a larva progresses from the first to the third instar, its size increases tenfold and, accordingly, its appetite also increases proportionately. It is when the maggots have reached the third larval instar that most of the corpse biomass is consumed and the loss of flesh is most pronounced. At the end of the third and final larval instar, the maggots go through a brief postfeeding, prepupal, wandering or migration (all synonymous) stage. At this time, they begin to crawl away from the corpse and have been known to migrate up to 50 meters from the food source and burrow into the soft soil or leaf litter. At the end of the migration stage, the last larval instar begins to shorten and turn darker with age, mobility is reduced, and formation of the pupal stage starts. It is the exoskeleton of the third larval instar that forms the outer hardened casing know as the puparium, and within this structure the pupa is

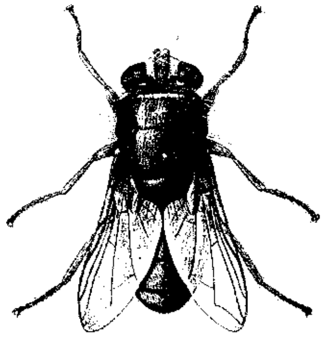


Figure 3 Adult blowfly, *Phaenicia sericata*. Reproduced with permission from Hall DG (1947) *The Blowflies of North America*. Entomological Society of America.

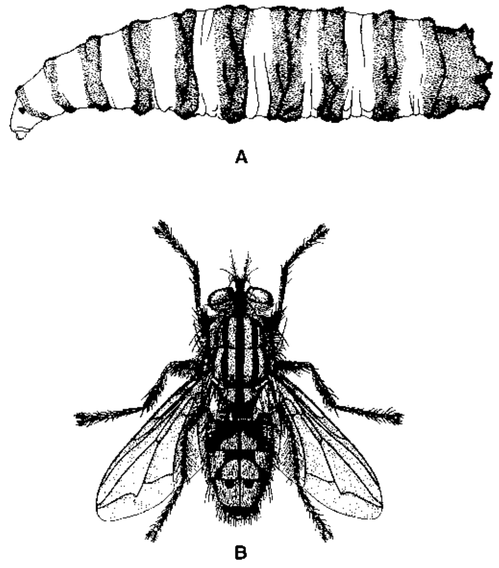


Figure 4 Flesh flies. **(A)** Larva; **(B)** adult. Reproduced from Gorman JR (ed.) (1987) *Insect and Mite Pests in Food: An Illustrated Key*. US Department of Agriculture, Handbook 655.

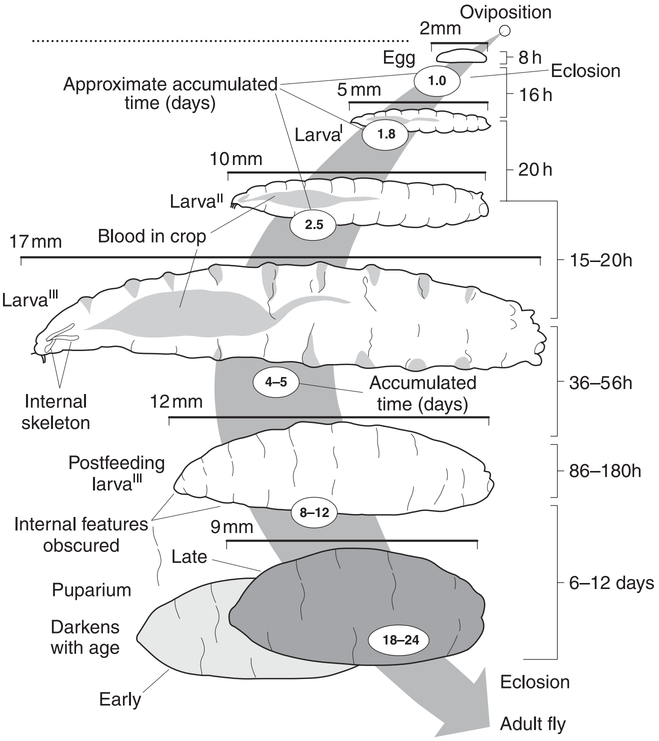


Figure 5 Hypothetical development of an immature blowfly. Reproduced with permission from Catts and Haskell (1990).

formed (Fig. 5). Inside each puparium a pupa slowly metamorphoses into an adult fly, which in several days time emerges to seek another corpse.

Importance of Temperature

Insects are poikilothermic (cold-blooded), in that their temperature is greatly influenced by the temperature of their macro- and microenvironment, and thus their development is closely tied to temperature. Generally, the warmer the temperature (up to some tolerance point of the insect), the faster development occurs (**Fig. 7**). The faster insects develop on a corpse, the faster decomposition will occur. A body left in the open on a hot summer day may be reduced to bones in a little over a week, whereas a corpse left out in the late autumn or winter may show little change for months. Entomologists have carefully studied the development of several species of carrion-feeding flies, determining their thermal requirements for completing each stage of development, from egg to adult (**Fig. 8**). These thermal requirements are expressed in temperature-time units called degree-days or degree-hours. It may be easiest to think of degree-days or degree-hours as accumulations of heat units over time. A hot summer day will accumulate abundant heat units, while a cool autumn day will accumulate few, if any. If a species of fly requires 400 degree-hours to reach the third larval instar, it may accumulate over 100 of those in a single,

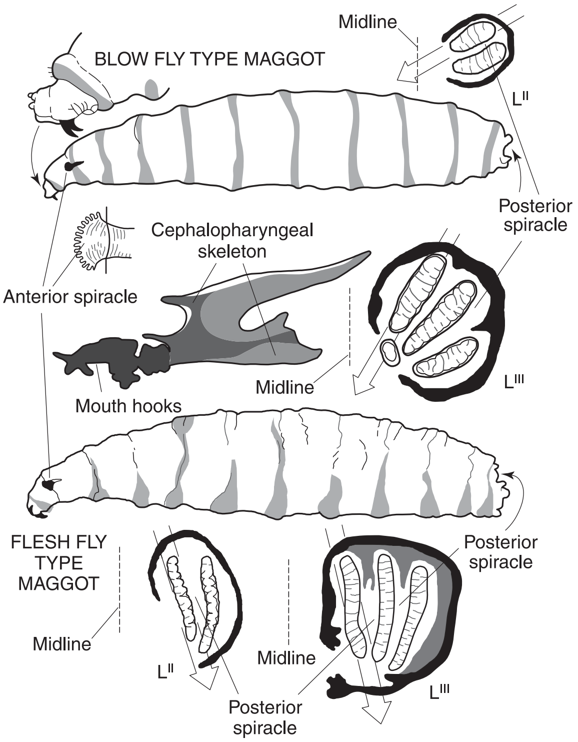


Figure 6 Immature blowfly (above) and flesh fly (below) showing posterior spiracles and internal structures of the head. Reproduced with permission from Catts and Haskell (1990).

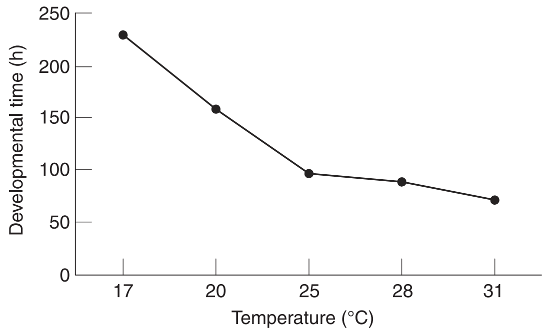


Figure 7 Relationship between blowfly development and temperature.

warm summer day, whereas it may accumulate only 10 on a cool autumn day. In addition, there is always some threshold temperature below which no development takes place, and below which adult flies are inactive. For most species, this threshold temperature is around 10°C. A corpse deposited at a time when the air temperature is below 10°C will generally not be colonized by flies until the temperature rises.

An exception to this temperature–developmental time relationship occurs when huge numbers of second and third larval instars feed in a concentrated 'maggot mass'. The activity of several hundred fly larvae in a small area can raise the temperature within this mass up to 30°C above the ambient air temperature, and fly larval development may proceed at a much faster pace than might be expected at a given temperature. If a maggot mass is present on a body, a forensic entomologist will usually record the temperature within the mass to help determine the PMI.

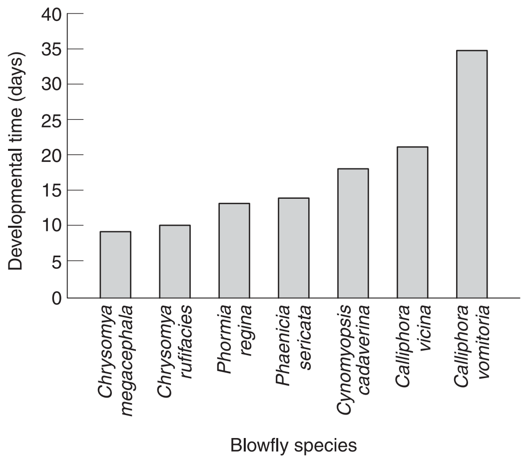


Figure 8 Developmental times for two tropical species (*Chrysomya megacephala* and *Chrysomya rufifacies*), two summer species (*Phormia regina* and *Phaenicia sericata*) and three cool-weather species (*Cynomyopsis cadaverina, Calliphora vicina* and *Calliphora vomitoria*). Both tropical species were reared at 28°C; summer and cool-weather species were reared at 26.5°C. Modified from Haskell *et al*. 1997.

Determination of the Postmortem Interval

A forensic entomologist can determine the postmortem interval by one of two methods. The first – and more precise – method involves identifying the species of flies (larvae and/or pupae) collected from the body at the time of discovery, determining the growth stage the flies are in, and calculating how long it would take them to reach this stage given the environmental temperatures at the scene. This tells the entomologist how old the flies are, in days. Because adult flies will usually lay eggs on a corpse within minutes or hours after death, knowing how old the fly larvae are gives a close approximation of time of death. In order to calculate the PMI accurately, the forensic entomologist must be able to identify the species of fly or flies present on the body, must be working with the oldest, and many times the largest, fly larvae present at the scene, must have access to nearby temperature records, and must know the thermal requirements for growth of the species of flies. Working backward in time from corpse discovery, an entomologist calculates the accumulated thermal units (degree-days or degree-hours) using local temperature records, which are usually obtained from the nearest weather station. The PMI is determined when the degree units calculated from local conditions match those drawn from baseline studies under controlled laboratory regimens, and are the ones required for the fly larvae to reach the developmental stage they were in when the corpse was discovered.

The second method of determining the PMI involves examining the composition of the insect community on the body and determining the successional stage of decay. This method is usually only employed in cases of advanced decay, after most of the flies have pupariated and emerged as adults. By this time most of the flesh has been removed from the corpse by maggots, so much of what remains is hair, skin and bone. Changes in the insect community and the corpse itself during this stage occur very slowly, making determination of the PMI much more difficult. The entomologist uses the presence of various beetle species and other insects to estimate the time of death, and usually must present a possible time range rather than an exact date.

Factors Affecting Determination of Time of Death and the Postmortem Interval

Any factor that makes a corpse inaccessible to insect colonization can affect the accuracy of determining time of death based on the PMI. As mentioned above, nighttime and cold temperatures, as well as rain, will

delay adult flies from ovipositing. In addition, factors that physically prevent insects from reaching a corpse can delay colonization for potentially long periods of time. Such factors include submergence under water, burial underground, storage in a sealed container, caskets or wrapping, and enclosure in a freezer, closed vehicle or building. While these factors will often delay insect colonization, rarely will they totally prevent it. Flies are very adept at locating corpses, and will usually find a way into a closed vehicle or building. A specific group of flies (Phoridae) have members that have been found 1–2 meters below the ground in association with coffined bodies, hence the name 'coffin flies'. Various studies have been conducted to determine how long the factors listed above will delay insect colonization of a corpse, and the forensic entomologist uses these to calculate accurately the time of death.

Collecting Entomological Evidence

When a corpse that contains insects is encountered, a trained forensic entomologist should be called upon to collect specimens. If a forensic entomologist is not available, it is imperative that a collection be taken in the field as soon as possible. If maggots are present on the body, the largest specimens must be collected, along with a representative sample of other sizes. The largest maggots will generally be the oldest, and thus represent the earliest colonizers of the corpse. If an entomologist is to make an accurate determination of the PMI, he or she must have possession of the oldest maggots present at the scene. The investigator should search the surface surrounding the body for maggots crawling away to pupate, and also dig in the surrounding soil (up to 2–3 meters away from the body) for fly puparia. Once again, failure to collect the oldest stage of developing flies will seriously undermine the determination of the PMI for the entomologist. If no maggots are visible, search the body – particularly around the eyes, ears, nose and mouth – for fly eggs, which will look like small grains of rice. Care should also be taken to collect insects from underneath the body once it has been removed. Detailed field notes as to the climatological conditions and location and condition of the body (e.g. rainfall, full sun, shade, fog, position of the body, wounds, etc.) and visual observations of the insects active in and around the remains should be made, as they can be of important value at a later time in the investigation. In addition, photographs of the crime scene should include close-ups of entomological evidence.

If flying insects are present over the body, an aerial sweep net can be used to collect fast flying or fast crawling adult insects. These can then be placed in a wide-mouth killing jar containing ethyl acetate and later pinned or placed in ethanol for subsequent identification.

The sample of insects collected from the body itself should be placed in a vial or jar containing alcohol (ethanol or isopropyl) along with a label (written in pencil) that includes the case number, date, time, collector and where on the body the specimens were collected. If at all possible, larvae first should be killed in boiling water and then transferred to ethanol to prevent shrinkage in larval size. However, this may not be practical, in which case they should be placed directly into alcohol. Maggots collected away from the body should be preserved separately, as should any fly puparia, and this information should be noted on the label. If possible, live specimens also should be collected. An entomologist may wish to rear these specimens to adults to aid in identification. Puparia can be placed in a jar with some of the surrounding soil, vermiculite or sand. A few live maggots can be placed in a jar and transferred as soon as possible to an aluminum foil envelope containing a small amount of fresh beef liver. Live specimens should be kept in a warm – not hot – place until handed over to an entomologist. All specimens, live and preserved, should be delivered to a qualified forensic entomologist as soon as possible for analysis. Copies of crime scene notes and photographs also should be provided.

Additional Value of Insects at Crime Scenes

While aiding in the determination of time of death by calculating the postmortem interval is the most common job of the forensic entomologist, analysis of insects at a crime scene also can reveal further information. Maggots and pupae can be analyzed for toxicological trace elements when, owing to decomposition, more traditional sources such as blood and tissue samples are no longer available. Traces of cocaine and other drugs have been obtained from insects present on a corpse, and have been used to determine probable cause of death. In addition, research has shown that maggots feeding on tissue containing cocaine experience accelerated growth. The presence of unusually large maggots, or 'super maggots' as they have been called, particularly those feeding around the nasal cavity, may indicate cocaine use by the deceased, and such maggots should be preserved for toxicological analysis.

The presence of maggot activity on areas of the

body other than the eyes, nose, ears and mouth prior to the decay stage of decomposition (Table 1) may indicate possible trauma sites. Adult flies will be attracted to wounds and will subsequently lay eggs in such areas. Maggot activity in the female urogenital and anal areas during the early stages of decomposition may be indicative of rape.

Entomological analysis also can be useful in nonhomicide cases, particularly those dealing with neglect and accidental death due to bites or stings. Neglect can often be demonstrated by the presence of maggots in living tissue, such as one might find in a nursing home for elderly people. Certain species of flies lay eggs only in living tissue, and the presence of their maggots on a body – even if the individual has subsequently died – point to invasion before death.

Deaths resulting from the bites or stings of insects, spiders and scorpions are fairly common. The marks left by these bites or stings are usually very small and may be overlooked. Death may be attributed to a heart attack rather than an accident, a conclusion that may be important for insurance purposes. A forensic entomologist, or a physician familiar with such cases, can often recognize evidence of bites and stings.

Summary

Forensic entomology is the study of arthropod-related evidence at a crime scene. Most often, forensic entomologists are called upon to assist in homicide investigations, usually to help in determining the time of death. Insects play an important role in the decomposition process, and colonization of a corpse follows a predictable succession of species and developmental patterns based on air temperatures. The most important group of insects that colonize a corpse is the true flies or Diptera, particularly blowflies. Blowfly larvae are responsible for consuming most of the decaying flesh, and the forensic entomologist uses known developmental times for each life stage of a particular fly species to determine the interval between time of death and corpse discovery (i.e. the postmortem interval). By examining the life stages present on the body at a crime scene, and using temperature data for the time period preceding corpse discovery, a trained forensic entomologist can calculate when the body was initially colonized by insects. Because colonization usually occurs shortly after death, this may be a close approximation of time of death. Factors that delay insect colonization, such as cold temperatures, darkness, rain, burial, etc., will affect the determination of the postmortem interval, and must be considered by the entomologist. Insect evidence can also be useful in providing toxicological and pathological information about the deceased.

See also: **Causes of Death**: Scene of Death; Postmortem Changes. **Drugs of Abuse**: Classification, including Commercial Drugs. **Toxicology**: Methods of Analysis – Postmortem; Interpretation of Results. **Time Since Death**.

Further Reading

Byrd JH and Castner JL (eds) (2000) *Forensic Entomology: The Utility of Arthropods in Legal Investigations*. Boca Raton, FL: CRC Press. LLC (in press)

Catts EP and Goff ML (1992) Forensic entomology in criminal investigations. *Annual Review of Entomology* **37**: 253–272.

Catts EP and Haskell NH (1990) *Entomology and Death: A Procedural Guide*. Clemson, SC: Joyce's Print Shop.

Goff ML (1993) Estimation of postmortem interval using arthropod development and successional patterns. *Forensic Science Review* **5**: 81–94.

Greenberg B (1991) Flies as forensic indicators. *Journal of Medical Entomology* **28**: 565–577.

Hall RD and Haskell NH (1995) Forensic entomology: applications in medicolegal investigations. In: Wecht C (ed.) *Forensic Sciences*. New York: Bender.

Haskell NH, Hall RD, Cervenka VJ and Clark MA (1997) On the body: insects' life stage presence and their postmortem artifacts. In: Haglund WD and Sorg MH (eds) *Forensic Taphonomy: The Postmortem Fate of Human Remains*, pp. 415–448. Boca Raton, FL: CRC Press.

Keh B (1985) Scope and applications of forensic entomology. *Annual Review of Entomology* **30**:137–154.

Rodriquez WC and Bass WM (1983) Insect activity and its relationship to decay rates of human cadavers in East Tennessee. *Journal of Forensic Sciences* **28**:423–432.

Smith DGV (1986) *A Manual of Forensic Entomology*. Ithaca, NY: Comstock University Press.

Ethanol *see* **Alcohol**: Blood; Body Fluids; Breath; Congener Analysis; Interpretation; Post Mortem.

ETHICS

R Weinstock, University of California, Los Angeles, and West Los Angeles Veterans Affairs Medical Center, Los Angeles, CA, USA
G B Leong, Center for Forensic Services, Western State Hospital, Tacoma, WA, USA
J A Silva, VA Outpatient Clinic, San Jose, CA, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0500

Introduction

Ethics in the forensic sciences is complex and challenging, as a result of functioning at the interface of science and law – two major disciplines with differing methods, values and goals. The law needs to obtain definitive answers in order to resolve disputes promptly and attain justice. In contrast, science reaches tentative conclusions subject to change with further evidence. Forensic science applies science to legal issues, but many differences exist between legal ethics and either scientific or professional ethics. There are specific ethical requirements for each scientific discipline with additional ethical requirements whenever scientific skills are applied to legal matters. Frequently, the two ethical requirements supplement each other. Many scientific disciplines facilitate forensic ethics by addressing the ethical aspects of the application of that discipline to legal issues, but not all disciplines do so. Whenever these requirements conflict, ethical dilemmas are created. Although there are many potential ethical problems in forensic science, most problems can be resolved by following Codes of Ethics or standards of good forensic practice.

Forensic Science Ethics and Personal Ethics

Forensic science ethics is the ethics of applying science to the law. Many ethical facets in forensic science are controversial. Some forensic scientists attempt to resolve ethical disputes by making an apparently arbitrary distinction between 'ethics' and 'morals'. However, these terms have been used interchangeably in philosophy for many years. The arbitrary distinction between 'ethics' and 'morals' enables those forensic scientists to avoid many ethical dilemmas by labeling certain ethical problems as only personal moral issues. Such ethical matters and dilemmas are thereby excluded from scientific and professional ethics discourse and consideration. The more appropriate distinction would be between personal ethics or morals, current professional and scientific ethics and the practitioner's view of what should be professional and scientific ethical behavior. Most issues can be considered from either a personal or a scientific and professional ethics perspective, even though complex ethical problems may not lead to a consensus.

In the personal sphere, 'ethics' or 'morals' refers to the concerns forensic scientists have that are based on personal ethical (moral) or religious considerations not derived from their professional and scientific roles. In the professional and scientific spheres, 'ethics' is the term traditionally used. 'Ethics' in this context refers to fundamental foundational rules or guidelines regulating conduct in scientific and professional disciplines or forensic science organizations. In contrast, rules of conduct or matters of etiquette are less fundamental matters that should be distinguished from ethics.

Organizational Forensic Science Ethics

For various reasons, not all professional and scientific organizations have ethical requirements or methods of enforcing them. When an organization does enforce its ethics, due process procedures are required both ethically and legally, in which accused members have an opportunity to hear the charges against them and present a defense. The advantage of ethics enforcement by a forensic science organization is that knowledgeable peers oversee one another. Otherwise such oversight may be taken over by outsiders, who may be insufficiently familiar with the discipline or profession to be able to be fair and effective.

Ethical requirements ideally should be enforceable. They should address minimally acceptable professional and scientific behavior. Behavior below these standards should result in a finding of an ethical violation with appropriate punitive sanctions. Sanctions can range from warnings and reprimands, to suspension and expulsion from an organization. A limitation on the ability of organizations to enforce ethical standards is that they cannot enforce ethical standards and requirements on nonmembers or on individuals who withdraw from membership. However, some organizations publish the fact that a member resigned from the organization while under an ethics investigation, in disciplines in which

licensing boards have authority, licenses could be suspended or revoked.

Although it is essential to meet basic minimal ethical requirements, the attainment of only a minimal threshold of ethical practice does not necessarily indicate good forensic practice. Failure to find an ethical violation is far from meaning that the professional or scientist necessarily is displaying impeccable ethics. Like the ‘not guilty’ adjudication in legal proceedings, it may mean only that there is an insufficient basis to prove that a practitioner has violated forensic science ethics.

Basic Minimal Ethics and Aspirational Ethics

Some organizations distinguish between ethical requirements representing basic minimal requirements, and aspirational standards or standards of good forensic science practice. Although many organizations confuse these two types of ethical standards, a clear distinction is important so that forensic scientists can determine which provisions represent basic minimal standards that must never be violated, and which represent a higher ideal threshold of desirable and exemplary conduct. The aspirational provisions represent desirable standards towards which forensic practitioners should strive. ‘Good’ forensic scientists will do so. In contrast to minimal standards, failure to achieve aspirational standards should not lead to sanctions. Sometimes, there could be an acceptable reason for not meeting the standards of good forensic practice.

An aspirational standard may sometimes be unenforceable because an assessment of the intent of the forensic scientist may be necessary in order to evaluate the behavior. Since intent is subjective, its determination may be difficult, if not impossible, to ascertain. Nonetheless, a good forensic scientist will try to meet the aspirational standard. Alternatively, a standard may be potentially enforceable, but may not be a minimal standard meriting sanction. Instead it might represent more a matter of desirable or good forensic practice.

Poor forensic practice does not necessarily mean unethical practice. Some inadequate forensic evaluations might represent lack of knowledge or failure to keep up to date in a forensic science discipline. The appropriate action in such instances might involve education rather than a punitive sanction. Other inadequate evaluations may result from time pressure and/or overwork that may contribute to overlooking some important aspects of a case. Although an evaluation may be inadequate, the negligence might not be of sufficient gravity to violate a basic minimal ethics requirement.

Codes of Ethics in Forensic Science Practice

Organizations like the American Academy of Forensic Sciences have developed a Code of Ethics and Conduct that the organization enforces. The Ethics Code states that members ‘shall refrain from providing any material misrepresentation of education, training, experience or area of expertise’. It also requires members to ‘refrain from providing any misrepresentation of data upon which an expert opinion or conclusion is based’. Additional provisions are part of the Code of Conduct. They are designed to prevent members from falsely claiming to represent an organization and from engaging in conduct adverse to the best interests and purposes of the organization. Such transgressions of a Code of Conduct may violate the rules of the organization but may not represent violations of fundamental forensic scientific ethics. The Code of Ethics and Conduct also has provisions describing due process procedures.

When enforcing a Code of Ethics, the following questions, as enumerated by Rosner, should be answered:

1. What provision is the member accused of violating?
2. What are the criteria for that provision?
3. What are the relevant data?
4. What is the reasoning by which it is determined that the member has or has not violated the specific provision?

It is important to write ethics provisions clearly in order to prevent ambiguity that could result in an ethics hearing degenerating into a popularity contest. For example, in the absence of clear guidelines, there is a danger that a verdict could primarily become dependent on whether a hearing panel likes an accused individual, whether the accused person’s views on a controversial issue are popular, or whether the accused is part of the ‘in’ group. As a result, the actual seriousness of the ethics offense would become much less relevant, with less effort expended towards determining exactly what happened and for what reason, thus obscuring the stated goal of the hearing.

If there are sections of the Code that potentially are unclear to a reader, clarifying interpretations should be disseminated to the organization’s membership. It is essential for members to have information in advance about the specific types of conduct that are prohibited (minimal standards), with a clear distinction from those that are encouraged (aspirational standards). If there is any potential ambiguity, a way must be found to provide the necessary clarification. One possible means is to print an explanation of the issue in the organization’s newsletter.

Standards for Good Forensic Practice

Aspirational standards are standards of good practice or even excellence. Good forensic practitioners should strive to attain these standards despite the fact that only minimal basic ethical standards are subject to enforcement. The American Academy of Forensic Sciences' Committee on Good Forensic Practice has developed the following Standards for Good Forensic Practice.

1. Forensic scientists generally should follow the standards of their respective disciplines. They should apply with care any assessment methods, technical skills, scientific and other areas of specialized knowledge, to legal issues and questions. They should always strive to do high quality work.
2. Forensic scientists should strive to keep current and maintain competence in their scientific disciplines. Although competence at minimum should be a goal, forensic scientists should strive for excellence.
3. Forensic scientists should demonstrate honesty and should strive for objectivity, by examining scientific questions from all reasonable perspectives and by actively seeking all relevant obtainable data that could distinguish between plausible alternative possibilities.
4. Forensic scientists should strive to be free from any conflicts of interest. They should possess an independence that would protect their objectivity. Any potential conflicts of interest should be disclosed. Work on relevant cases should be avoided or discontinued if objectivity may be compromised.
5. Forensic scientists should undertake cases and give opinions only in their areas of expertise, attained through education, training and experience.
6. Forensic scientists should attempt to identify, deter and help eliminate unethical behavior by other forensic scientists through methods such as discussion with a colleague, education, and if unsuccessful, by filing an ethics complaint.
7. It is essential to recognize that honest differences of opinion exist and do not imply unethical behavior by either expert. The legal adversary system includes opposing attorneys seeking out experts with favorable opinions. Forensic scientists should not be blamed unfairly for unpopular verdicts, honest differences of opinion, or the vagaries of the legal system.
8. Passions against an opposing disagreeing expert, or personal animosity, should not constitute the basis for an ethics complaint. Ethics complaints must be made in good faith. If based on passion alone, such ethics complaints themselves are inappropriate.
9. Forensic scientists should present their opinions to the trier of fact in concise understandable language, but care must be taken since such efforts can result in oversimplification and loss of some precision. In their efforts to communicate effectively, forensic scientists should strive to be as accurate as possible and avoid distortion. Every reasonable effort should be made to ensure that others (including attorneys) do not distort the forensic scientist's opinions.
10. Forensic scientists should strive to instill the highest ethical and scientific standards in their students and colleagues through such means as teaching, supervision, setting a good example, publications, and presentations at meetings.
11. Forensic scientists should strive for excellence and the highest degree of integrity. Forensic opinions should not be based on undisciplined bias, personal advantage, or a desire to please an employer or an attorney.
12. When forensic scientists are asked to express opinions on a legal issue, they should make every effort to become familiar with the applicable legal criteria in the pertinent jurisdiction. They should take care to reach only those legal conclusions that result from proper application of the data to that legal issue.
13. Unlike attorneys, forensic scientists are not adversaries. They take an oath in court to tell the whole truth. They should make every effort to uphold that oath.
14. When a forensic scientist accepts any privileged information from an attorney, care should be taken to ensure that all such information is kept confidential and does not reach the opposing side. After accepting such information, forensic scientists should not provide their services to the opposing side unless legally ordered to do so. Forensic scientists should alert attorneys not to make payment or provide privileged information, if they wish to retain the option to be employed by the opposing side.

Ethical Problems in the Forensic Sciences

Some forensic scientists confuse their role with that of an attorney. The attorney's role is to present the best one-sided case for his/her client, the only exception being not to argue for anything the attorney knows to be untrue. Unlike an expert witness, attorneys take no oath in court. In contrast, forensic scientists take an oath 'to tell the truth, the whole truth, and nothing but

the truth'. If forensic scientists assume the attorney's total advocacy role and present the best one-sided case for the retaining attorney regardless of the 'truth' and the oath taken, the forensic experts may be perceived as 'hired guns'. They thereby vilify not only their own reputation, but also taint that of the entire field. Many forensic scientists consider the 'hired gun' problem the most serious ethical problem in forensic science.

A 'hired gun' can establish a track record of testifying for both the plaintiff (prosecution) and defense, depending on the side that hires the expert. An effective 'hired gun' can appear impartial and objective by giving one-sided misleading persuasive explanations for whichever side hires him/her for a particular case. 'Hired guns' not only always make the best case for the side hiring them, but alternatively they also could do so for a side for which they have a personal bias, regardless of which side they actually believe is right. In contrast, it is possible to have a bias and preferentially work for one side, yet refuse to take on a case unless the forensic scientist honestly believes an opinion is true with the greatest objectivity possible. The distinction between a hired gun and an expert with an idiosyncratic opinion is not always clear-cut without knowing what is in the mind of the expert. Even if experts firmly believe an opinion because of personal bias so that they honestly may be telling what they believe to be true, others frequently might incorrectly view the opinion as a dishonest distortion of the truth. Sometimes, the problem is an insufficient effort by the expert to be objective, or an honest difference of opinion between forensic scientists.

Honest differences of opinion do exist as they do in every science and profession including law, and it is unfair to blame forensic scientists and forensic science for the battle of the 'experts'. Additionally, once a forensic scientist reaches an honest objective opinion, the pressures of our adversary system encourage experts to advocate for their opinion. It is difficult to remain totally objective when defending an opinion. Subtle pressures can also bias an expert, such as the wish to please an attorney or an employer. An honest expert should make every effort to form opinions based on the evidence even if contrary to the wishes of the hiring person or institution. There are also questions about what it means to tell the truth, the whole truth and nothing but the truth. Does it mean the expert should try to offer the whole truth and not a misleading part of the truth to the degree that the attorneys and the court will permit? Or is it enough to present only that part of the truth that will help the case and rely on good cross-examination (which may not happen especially if the case is settled before trial) to bring out the rest of the truth? Or should some portions of the truth that may lead to conclusions contrary to the expert's opinion be presented solely as a tactical maneuver to make a preemptive strike so that those aspects can be debunked before the other side has a chance to present them?

Should a forensic scientist try to participate in the legal system only in ways that further justice or should the forensic scientist solely answer the questions asked and trust the adversarial system to usually achieve justice? The adversarial approach exists in those countries like the United States that base their system on English common law. Should forensic scientists lend their expertise to sides trying to achieve an 'unjust' outcome? Is it presumptuous for forensic scientists to try to further only what they see as justice or what is consistent with their professional scientific or personal values? Answers to these questions presently lack a consensus and a resolution does not seem likely in the foreseeable future.

Foundations of Ethics

Are ethical guidelines arbitrary rules, or are they based on something more basic and fundamental? Ethical codes address 'important' issues, and seem to represent more than codes of conduct or codes of etiquette. Sometimes codes confuse etiquette (such as not criticizing another scientist or professional or not hurting an organization) with fundamental ethical issues. There are questions whether any ethical truth is objectively right or wrong, or whether all ethics is subjective and one code is as good as another. Some, who view ethics as subjective, see it as solely dependent on the mores of a specific culture or scientific discipline. They believe that what is considered ethical is only what feels morally right to individuals in that culture or group.

Some consider that the only foundation for ethics and morals is religion. However, religion itself lacks objectivity not only within the confines of a single religion, but also because there would be no consensus as to which of many different religions should be the standard bearer. Some religions in the past have advocated human sacrifice. Did that make it 'right' for them? There is a long tradition of secular ethics independent of religion in western democratic countries. Although religion can be a basis for ethics, many religiously observant people can be unethical and do 'wrong' things despite engaging in religious practice. Atheists can be very ethical and committed to philosophical study to determine what is 'right'. Although religion can undoubtedly motivate many people to do what is 'right', by no means is religion a necessary foundation for ethics and morals.

Two primary schools of thought exist regarding the basis for ethics. One is consequentialist and proposes

that the ethical thing is whatever leads to the best consequences, such as the most good or happiness for the most people. A subset of this position is utilitarianism, insofar as what is most useful is ethical. Another school is deontological and bases ethics on an intrinsic duty. Actions are intrinsically right or wrong regardless of consequences. An example of the latter is the philosopher Immanuel Kant's categorical imperative – to do what you believe any ethical person in your position should do, such that the maxim you are following should be a universal principle for all to follow.

There can be problems with following any of these positions too rigidly. You might consider justice an intrinsic deontological duty, much as Kant did. Does that mean though that every crime no matter how small should be investigated until the perpetrator is punished? It takes consequentialist considerations to balance this duty with the desirability of spending government money on other things that might be more important for society's welfare. Investigating every crime thoroughly until the criminal is caught would use up all our society's resources. No funds would be left for other desirable purposes. Additionally, if deontological considerations were primary, two deontological duties might conflict without any principle to resolve the conflict.

Similarly, following only consequentialist considerations results in problems. For example, it might decrease crime in society if it were perceived that every criminal was always swiftly caught and punished and no criminal ever got away with a crime. To achieve this end it might help to choose individuals from an unpopular group such as prior offenders, try them in secret, find them guilty, and execute them quickly and publicly. Such a procedure might in fact have some deterrent value and decrease crime especially if it were used for white-collar crimes like embezzling money or income tax fraud. However, deontological considerations of justice convince us that there is something wrong with this procedure. There are even contrary consequentialist considerations. Government would become very powerful, and any of us could be arbitrarily tried, convicted, and executed even if we were innocent. All of us would be at potential risk. Thus, a purely consequentialist foundation for ethics would also provide no way to decide between these competing consequences to determine the right action.

The best approach requires a consideration of both types of ethical foundations, but there is no higher order rule to help us in balancing these considerations. The theoretical approach to resolving ethical dilemmas essentially uses universal ethical principles as 'axioms' from which particular ethical judgments are deduced as 'theorems'. A solution to a problem in a particular case is deduced from the ethical principle. However, there usually is no higher order ethical rule or principle to help us decide among competing ethical principles.

In contrast, the practical approach, known philosophically as 'casuistry', does not involve universal theoretical principles and rules but requires us to consider the issues from the perspective of a particular case. General ethical principles and rules essentially are 'maxims' that can be fully understood only in terms of paradigmatic cases that define their meaning and force. However, there can be and often are differences of opinion about which facets of a case should be overriding, with no ethical guidance to help us in weighing and balancing competing ethical maxims.

In both instances, deontological and consequential considerations can assist us in determining what is 'right'. However, when ethical rules and principles conflict, there generally is no rule, principle, or maxim to help us resolve the conflict. Therefore, it is essential that forensic scientists are able and prepared to think through and resolve ethical dilemmas themselves. In difficult situations, more than one solution usually should be considered acceptable, and ethics committees should not place sanctions on forensic scientists for following either course. In order to help reach the 'best' solution in such cases, consultation should often be obtained from experienced forensic scientists knowledgeable about ethics issues.

Ethical Dilemmas

Ethical dilemmas often defy consensus solutions. When ethical responsibilities conflict, some forensic scientists always give the legal needs priority. The majority, however, usually give the legal needs priority, but make exceptions if the law desires an action that violates serious aspects of the values and ethics of their own scientific discipline or profession. In such situations, their professional or scientific values, ethics, and responsibilities may outweigh the desires of the legal system, and preclude their involvement in that aspect of a legal case. Personal ethics and morals can also preclude such involvement. Although the legal system can establish legal ethics, only scientific, professional, and forensic science disciplines and organizations can establish forensic science ethics. The law cannot determine what is ethical in forensic science; it only can determine what is legal.

Ethical dilemmas occur when there are conflicting ethical duties. An example of such a conflict occurs in forensic psychiatry, in circumstances in which a practitioner might sometimes believe that the most accurate and truthful assessment would be attained by misleading the person being evaluated. Such deception

is often legal, but it violates forensic psychiatric ethics as promulgated by the American Academy of Psychiatry and the Law. Most forensic psychiatrists follow the guideline of the forensic psychiatric profession regardless of the frequent legality of such deception. In another instance, many consider the presentation of aggravating circumstances in a death penalty phase of a trial a violation of medical ethics, even if such aggravating circumstances are true and such testimony in the United States violates no current ethical requirement. The same is true if individuals incompetent to be executed are treated to make them competent. These problems occur when medical or psychological skills are used to evaluate an individual for forensic purposes.

The specter of ethical conflict, however, is by no means limited to forensic psychiatry and psychology or other branches of forensic medicine, but it occurs in all forensic sciences. The use of DNA evidence in the courtroom, although seemingly based on sophisticated modern scientific techniques, with unusually high levels of reliability and certainty, is another example of a highly controversial area. Problems in other aspects such as data collection can be questioned. In many ways it seems counterintuitive that there can be such divergence in professional opinions when accurate data such as DNA evidence are introduced. Nonetheless, controversy does exist about the interpretation of DNA data among highly respected forensic scientists, such as in the widely publicized criminal trial of O.J. Simpson. Does that necessarily mean that one side is unethical or is it possible for honest forensic scientists to hold strong opposing views? One possible explanation is that the highly accurate nature of DNA evidence has made many problems more significant. Potential ethical and data collection problems that had been readily ignored when identification methods were less accurate suddenly have begun to loom in importance. Many such controversies exist in all forensic science disciplines.

Conclusion

Although the ethical course of action is clear in the majority of situations, it is essential for the 'good' forensic scientist to be knowledgeable about ethics in order to be able to resolve ethical dilemmas when difficult situations arise. It is essential to know the minimal requirements in a Code of Ethics in order to stay out of trouble and avoid sanctions. However, such Codes do not and cannot address all contingencies since differing requirements as well as aspirational standards may conflict. Generally, there is no higher order rule telling us how to resolve such a conflict. In such instances, sometimes with knowledgeable help, forensic scientists must work out their own ethical solutions. Such dilemmas are likely to occur when different disciplines with differing ethics and values intersect, like science and law, in forensic science practice.

Various organizations address different issues in their codes. Two minimal standards for forensic scientists are the need to be clear and truthful, and not to distort credentials and data. However, the good forensic scientist should strive for more than only staying out of trouble and meeting minimum standards. 'Good' forensic scientists should strive for excellence both in their ethics and in their forensic science work.

Ethical complexity in the forensic sciences is not a reason to avoid forensic practice. Forensic scientists should accept the challenge, but should be prepared to confront and assess potential ethical conflicts. Myriad problems beset all disciplines, especially those that interface with two very different disciplines like science and law. Forensic science is an interesting, stimulating, and productive vital occupation, but it is necessary to become informed about as many facets as possible in order to become a good ethical practitioner. In most cases, despite the potential ethical dilemmas that have been enumerated, the forensic science Codes of Ethics provide the required minimal ethics solutions. Those wishing to be good forensic science practitioners who strive for excellence, generally can find guidance in standards of good forensic practice. It is only in relatively rare cases that ethical standards conflict or there are conflicts between provisions in the standards of good forensic practice. In such instances, forensic scientists should be prepared, perhaps with consultation, to perform their own ethical analyses.

See also: **Basic Principles of Forensic Science**. **Legal Aspects of Forensic Science**.

Further Reading

AAFS Committee on Good Forensic Practice (1999) Advisory opinion. *Academy News* **20**:24.

American Academy of Forensic Sciences (1998) *Code of Ethics and Conduct*. Colorado Springs: American Academy of Forensic Sciences.

American Academy of Psychiatry and the Law (1998) *Ethical Guidelines for the Practice of Forensic Psychiatry*. Bloomfield, CT: American Academy of Psychiatry and the Law.

Appelbaum PS (1997) A theory of ethics in forensic psychiatry. *Bulletin of the American Academy of Psychiatry and the Law* **25**:233–247.

Beauchamp TL (1982) *Philosophical Ethics: An Introduction to Moral Philosophy*. New York: McGraw Hill.

Beauchamp TL and Childress JF (1994) *Principles of Biomedical Ethics*, 4th edn. New York: Oxford University Press.

Callahan JC (1988) *Ethical Issues in Professional Life*. New York: Oxford University Press.

Code of Ethics for the Academy Editorial (1986) *Journal of Forensic Science* **31**:798–799.

Diamond BL (1992) The forensic psychiatrist: consultant vs. activist in legal doctrine. *Bulletin of the American Academy of Psychiatry and the Law* **20**:119–132.

Diamond BL (1994) The fallacy of the impartial expert. In: Quen J (ed.) *Selected Papers of Bernard L. Diamond*. Hillside, NJ: Analytic Press.

Dyer AR (1988) *Ethics and Psychiatry*. Washington, DC: American Psychiatric Press.

Hundert EM (1987) A model for ethical problem solving in medicine with practical applications. *American Journal of Psychiatry* **144**:839–846.

Jasinoff S (1995) *Science at the Bar: Law Science and Technology in America*. Cambridge: Harvard University Press.

Jonsen AR and Toulman S (1988) *The Abuse of Casuistry*. Berkeley, CA: University of California Press.

Katz J (1992) 'The fallacy of the impartial expert' revisited. *Bulletin of the American Academy of Psychiatry and the Law* **23**:141–152.

Mills DH (1997) Comments from the perspective of the AAFS ethics committee. *Journal of Forensic Science* **42**:1207–1208.

Murphy J and Coleman JL (1990) *Philosophy of Law: An Introduction to Jurisprudence* (revised edition). Boulder, CO: Westview Press Co.

Nordby JJ (1997) A member of the Roy Rogers Riders Club is expected to follow the rules. *Journal of Forensic Science* **42**:1195–1197.

Rachels J (1993) *The Elements of Moral Philosophy*, 2nd edn. New York: McGraw-Hill.

Rosner R (1994) *Principles and Practice of Forensic Psychiatry*. New York: Chapman & Hall.

Rosner R (1997) Foundations of ethical practice in the forensic sciences. *Journal of Forensic Science* **42**:1191–1194.

Rosner R and Weinstock R (1990) *Ethical Practice in Psychiatry and the Law*. New York: Plenum Press.

Stone AA (1984) *Law, Psychiatry and Morality*. Washington, DC: American Psychiatric Press.

Weinstock R, Leong GB and Silva JA (1991) Opinions expressed by AAPL forensic psychiatrists on controversial ethical guidelines: a survey. *Bulletin of the American Academy of Psychiatry and the Law* **19**:237–248. Erratum **19**:393.

Weinstock R (1997) Ethical practice in the forensic sciences – an introduction. *Journal of Forensic Science* **42**:1189–1190.

Weinstock R (1998) Letter to the editor. *Journal of the American Academy of Psychiatry and the Law* **26**:151–155.

Ethnicity *see* **Anthropology**: Determination of Racial Affinity.

Ethyl Alcohol *see* **Alcohol**: Blood; Body Fluids; Breath; Congener Analysis; Interpretation; Post Mortem.

EVIDENCE

Contents

Classification
Statistical Interpretation of Evidence/Bayesian Analysis

Classification

I Freckelton, Owen Dixon Chambers, Melbourne, Victoria, Australia

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0501

Introduction

Evidence comes before the courts and tribunals in many different forms. Sometimes it is determined to be pertinent to issues in dispute; at other times it is found not to be germane at all. Then it is said not to be 'relevant'. On some occasions it is peripheral to the

resolution of a case. On other occasions it is central. It can constitute 'direct evidence' of a fact in dispute but it can also form part of a collection of matters which tend to indicate a conclusion advantageous to one party in the litigation and so is 'circumstantial evidence'. Sometimes it is 'reliable', although there are a number of different criteria that can be applied to evaluate reliability. Sometimes it bears the hallmarks of being spurious. Sometimes it is 'hearsay', by reason of coming from a source that cannot be made readily accountable, by reason of its being outside the purview of the court or tribunal, so it is of questionable value.

Evidence in the form of 'tendency', 'propensity' and 'coincidence' seeks to induce an inference that there is consistency between previous acts and those charged, or between the perpetrators of such acts and the persons charged. Its admissibility and its utility as evidence depends on a number of important preconditions.

Evidence can be in the form of actual objects and so can be classified as 'physical' or 'real'; it can be in the form of a demonstration or experiment that the tribunal of fact can observe ('demonstrative evidence'); and it can be in the form of words recounting things that have been done and inferences that have been drawn ('parole evidence').

Some evidence is 'lay' evidence in that it is not given by an expert about their area of expertise. Other evidence is given by a person possessed of specialized knowledge by reason of their skill, training or experience. This is 'expert evidence'. Evidence by lay persons or experts can be given in the form of evidence of 'fact' – what a person did, heard, saw or felt – or it may be given in the form of 'opinion' – what a person concluded by inference from data.

Evidence that can be associated with a common source with a high degree of probability, such as matching fingerprints or handwriting, is often described as evidence exhibiting 'individual characteristics'. By contrast, evidence exhibiting 'class characteristics' is found when the evidence can be associated only with a class but not with a single source. This occurs, for instance, with the matching of samples of 'Type AB' blood, unless DNA profiling is undertaken on it, a form of test that can elicit individualization.

Relevant Evidence

Evidence is only permitted to be adduced before courts if it is 'relevant'. Relevance exists where the proposed evidence has a tendency to make the existence of any fact that is of consequence to the determination of a criminal or civil action more or less probable than it would be without the evidence. Another formulation of relevant evidence is evidence that, if it were accepted, could rationally affect (directly or indirectly) the assessment of the probability of the existence of a fact in issue in the proceeding.

The distinction between relevant and irrelevant evidence is fundamental to the use to which evidence in any court case can be put. The distinction determines its admissibility and also its probative value. For this purpose, 'probative value' can be regarded as the extent to which the evidence could rationally affect the assessment of the probability of the existence of a fact in issue. Relevance is context-dependent and so judges' findings about relevance are of minimal significance in terms of establishing precedent. However, a determination that evidence is not relevant, for instance because it does not constitute evidence that could assist the trier of fact in making a particular finding, is important because it results in the removal of such information from the material that goes before the trier of fact. Evidence determined by a judge not to be relevant is inadmissible.

Direct and Circumstantial Evidence

Evidence that tends to prove facts that must be proved in a case, for instance in a criminal prosecution, can be of two kinds: 'direct' and 'circumstantial'. A litigant on occasion can adduce evidence which, if accepted, would prove a fact in issue without the need for any inference to be drawn from one fact to another. This is 'direct evidence'. If the question is whether a person is dead and the person's dead body is produced, that is direct proof of death. Where the court employs its own senses to evaluate such evidence, it is 'direct real evidence'; where the court also has cause to depend on its assessment of the veracity of a witness, it is 'direct testimonial evidence'.

By contrast, if evidence is adduced that the person suspected of being deceased disappeared mysteriously at the seaside 2 years previously after leaving a suicide note together with detailed arrangements for the resolution of his personal affairs, this constitutes 'circumstantial', or 'presumptive', evidence of his death. Lord Chancellor Cairns in *Belhaven and Stanton Perlerage* (1875) 1 App Cas 278 at 279 summed up the functioning of circumstantial evidence well: 'You may have a ray of light so feeble that by itself it will do little to elucidate a dark corner. But on the other hand, you will have a number of rays, each of them insufficient, but all converging and brought to bear upon the same point, and when united, producing a body of illumination which will clear away the darkness which you are endeavouring to dispel.'

Although, in general, direct evidence is more

reliable than circumstantial evidence, direct evidence is subject to a range of factors that can reduce its reliability, including mendacity of witnesses, and the potential for perceptual and interpretative errors. However, the stuff of ordinary human experience, as well as of forensic evaluation, is the drawing of inferences from data in order to assess the likelihood of hypotheses. Circumstantial evidence in some contexts can therefore have considerable utility and reliability for the fact-finding process.

Real Evidence

A fact may be proved by real evidence, cognisable by the sense of a court and not reported through the testimony of a witness. It can take many forms and is often received if of sufficient relevance and reliability. Thus insanity may be proved by demonstration of derangement on the part of the accused; pregnancy can be proved by the appearance of an appropriately swollen belly; the fact that a child is under the age of 16 may be proved by how the child looks; the appearance of a person may be evidence of their nationality. In an old murder case a court was able to conclude by inspection of the positions of the death wound and of the hole in the waistcoat of the deceased that 'the wound given by the pistol under the right pap could no way happen by any position of the pistol in the bosom of the deceased, by the pistol going off by itself.' *R* v. *Reason* (1722) 16 How St Tr 42.

One form of real evidence is documentary evidence in which documents are tendered to the court as both proof of their existence and on occasions of what is written within them.

A further category of real evidence is the conduct of scientific or practical experiments in the presence of the jury. However, evidence of such matters is often circumscribed by trial judges lest it arouse undue prejudice in the minds of jurors, such as the sight of particularly gruesome or distressing human or other remains. The admissibility of real evidence is also limited by reference to the extent to which items such as photographs accurately represent the facts, are fair and lack a propensity to mislead and can be verified on oath by a person capable of doing so.

Jurors are also often cautioned against themselves functioning as experts in interpreting DNA autoradiographs, comparing tool marks or evaluating similarities in handwriting or fingerprints. The line they are permitted to tread is a thin (and unclear) one with jurors entitled themselves to examine exhibits and draw their own inferences, so long as they do not substitute their own judgments for those of expert evidence before them.

A category of real evidence is demonstrative evidence whose purpose is to explain or illustrate. Computer-generated re-enactments of matters such as aeroplane or motor vehicle collisions, or of failures of equipment or even of criminal assaults, are examples of such evidence. In general, an adequate foundation must be laid to establish both the authenticity of such evidence and its relevance. This is done in relation to re-enactment evidence by showing that the underlying data are accurate; proving that the process by which the data were entered into the computer provides reasonable assurance that error was avoided; and demonstrating that tests were used to maintain the accuracy and reliability of the relevant hardware and software. In addition, in many jurisdictions it must be shown that such evidence is more probative than it is prejudicial before it is admitted.

Parole Evidence

Most evidence in the courts is 'parole' or oral evidence in the form of testimony. It is often called 'testimonial evidence'. Such evidence is generally given on oath and from memory, save where the witness is permitted to refer to notes or other documents. The 'parole evidence rule' precluded oral evidence, once an agreement had been reduced to writing, about what passed between the parties either before or after a written instrument (such as a contract) was created so as to subtract from, add to or vary what was in the written document. However, it is subject to many qualifications in all jurisdictions.

Oral testimony is the mainstay of evidence in the Anglo-American tradition, allowing decision-makers to observe the demeanor and presentation of witnesses and thereby to gauge their credibility and veracity. Social science research has demonstrated that such evaluations can be flawed by reason of false assumptions and attributions. However, findings as to credibility are the bailiwick of the trier of fact and are amenable to appeal only to a limited extent.

Class and Individual Evidence

If it is shown that there is a match between an evidentiary exhibit found at the scene of a crime and an exhibit found in the possession of a person accused of the crime, the question then arises as to the significance of such a match. It arises in the context of many forms of matches, for example, fingerprints, DNA profiles, glass-refractive indexes, paint spicules, hairs, fibers, shoe impressions. Some evidence possesses highly identifying features. It is termed evidence that possesses 'individual characteristics'. Examples include DNA profiling evidence, together with dactylography or fingerprinting evidence, and a

range of forms of comparison evidence, such as toolmark, handwriting, tire or footwear impression evidence and similar.

In general terms, the value of a piece of individual evidence is inversely proportional to the chance of false association. This chance depends on the number of comparisons which are attempted – usually against data within a pertinent database. The larger a database, the more significant a match becomes and the greater the potential for contrast between individual evidence and evidence relating to class.

Thus evidence which establishes that an accused person had soil of a certain kind on her shoes and that the deceased was murdered in an area where that soil is to be found may be admissible against the accused. It is evidence that cannot uniquely identify an accused, so it can be said to possess 'class characteristics'. It may, nonetheless, constitute an important part of the evidence against her. However, for the trier of fact to evaluate the probative value of such evidence, and for the trial judge to determine whether such evidence is more probative than it is prejudicial, it is necessary for evidence to be adduced about a range of factors which would enable the trier to evaluate the significance of such a 'match'. Included among these is the ubiquity of such soil within the relevant geographical area, thereby enabling the trier of fact to calculate how likely it is that a person with no involvement in the homicide would have such soil samples on their shoes.

Reliable Evidence

Fact finders need evidence on which they can place reliance, evidence that is sound and in which trust can be reposed in determining whether an assertion in a civil or criminal case is established to the necessary degree of proof. In some jurisdictions, reliability of expert evidence is a precondition to its being admitted. This has shone light on the meaning of 'reliability' as a forensic concept.

In most jurisdictions evidence can be excluded, at least in criminal cases, where it is accounted more prejudicial than probative. Reliability is a factor in determining probative value.

In the United States, Canada, New Zealand and to some degree in Australia reliability is a determinant of the admissibility of expert evidence. In England it is not. In the United States, the influential Supreme Court decision of *Daubert* v. *Merrell Dow Pharmaceuticals*, 125 L Ed (2d) 469; 113 S Ct 2786 (1993) prescribed four indicia of reliability: (1) whether scientific evidence can be or has been tested, namely its falsifiability, refutability or testability; (2) whether the theory or technique has been subjected to peer review and publication as a means of increasing the likelihood that substantive flaws in methodology will be detected; (3) the known or potential rate of error and the existence and maintenance of standards controlling the technique's operation; and (4) whether a technique has gained general acceptance within the scientific community. The test, therefore, evaluates reliability of opinion evidence by a combination of scientific analysis and deference to the views of its legitimacy within the relevant intellectual marketplace. In other jurisdictions, extensive lists of indicia of reliability have been developed. They possess varying degrees of specificity in terms of the integrity required within the scientific process and the criteria for gauging it.

Hearsay Evidence

Evidence of a statement made to a witness by a person who is not called as a witness may or may not be hearsay. It is hearsay and frequently inadmissible when the object of the evidence is to establish the truth of what is contained in the statement. It is not hearsay and tends to be admissible when it is proposed to establish by the evidence, not the truth of the statement, but the fact that it was made. The preclusion upon evidence that is hearsay arises out of the need by the courts to rely on evidence in which confidence can be reposed. If evidence is at second or third hand and is not capable of effective evaluation because of the absence of its primary source, the evidence is of its nature unreliable. Hence the preclusion in most jurisdictions on much evidence that falls within the category of hearsay. However, the rule against hearsay evidence is subject to a number of qualifications and exceptions, and in some jurisdictions, such as the federal jurisdiction in Australia, has been significantly attenuated.

Tendency, Coincidence and Propensity Evidence

'Tendency evidence' or 'propensity evidence' is evidence of conduct adduced to prove a person or institution's tendency to act in a particular way. Attempts are frequently made to adduce such evidence to invite the inference of conforming conduct. This is 'propensity reasoning'. 'Propensity evidence' often takes the form of evidence that an accused person has committed wrongful acts similar to those with which he or she is now charged. The inference is then invited that because the accused has a propensity to commit a kind of crime of the sort charged, he or she has done so on this occasion. Such evidence can

go to identification of an accused, if there is evidence of an unusual criminal propensity; it may rebut a defense of accident and it may constitute evidence of 'relationship', 'association' or 'passion'. The tests for the admissibility of tendency or propensity evidence differ from jurisdiction to jurisdiction. However, such evidence is more probative, the more similar the past conduct and the charged conduct and the more unusual such conduct. However, such evidence is highly prejudicial and is often determined to be inadmissible or of little assistance because of being insufficiently helpful to the trier of fact.

'Coincidence evidence' is evidence that relies on the improbability of two or more events occurring coincidentally, namely totally independently of each other. The more strikingly similar the events, such as crimes, and the less common, and the more unrelated by collateral circumstances such crimes, the more compelling such evidence is.

Expert and Lay Evidence

Evidence is also classified by reference to the expertise of the witness. Most witnesses are lay witnesses and not permitted to give evidence in the form of opinions, except as shorthand expressions of fact (see below).

Some witnesses, who possess specialized knowledge by reason of their skill, training or experience are designated 'expert witnesses'. However, this is a flexible and often uncertain designation as all experts emerge gradually in knowledge and in status from the lay toward the expert. Such a transition is a function of a series of factors including extent of knowledge, practical exposure to an area of experience, currency of expertise, and degree of specificity of expertise. With increasing subspecialization in most disciplines, it is increasingly clear that mere possession of broad academic qualifications or membership of a learned society is not enough to designate a person an expert for forensic purposes. However, in most jurisdictions the provenance of expertise is not important – it is the actual possession of the pertinent expertise that matters, not how it was acquired.

Fact and Opinion Evidence

The status of expert, as against lay, witness is important in determining whether or not a witness is entitled to proffer opinions to a court. Given the potential for expert witnesses to be especially influential because of their knowledge, articulateness and testimonial experience, the designation as expert is of consequence. In general, lay witnesses are confined to evidence of what they have done, seen, heard and felt. They can only offer opinions which are shorthand expressions of fact, such as that a person appeared 'sad' or that a vehicle was traveling 'fast'.

The privilege of expressing opinions, of drawing inferences from facts, is that of the expert witness. However, the distinction between facts and opinions is not always easy. As long ago as 1898 Thayer wrote that, 'In a sense all testimony as to matters of fact is opinion evidence; i.e. it is a conclusion from phenomena and mental impression.' As most language embodies inferences of some kind, it is not possible wholly to dissociate statement of opinion from statement of fact. All statements are in some measure inferences from experience. Judge Learned Hand accepted the realities of the arbitrariness of the dichotomy frankly in 1926: 'the line between opinion and fact is at best only one of degree, and ought to depend thoroughly upon practical considerations as for example the saving of time and the mentality of the witness.' (*Central R New Jersey* v. *Monahan*, 11 F 2d 212 (1926)).

The distinction between fact and opinion evidence is at its most significant for legal purposes where the expertise of a witness and so their entitlement to give opinion evidence are in issue. For instance, in two High Court decisions in Australia evidence was allowed to be adduced from people with practical experience in, but no academic background in studying, the jackknifing of semitrailers and in the handling of semiarticulated vehicles on a certain road in certain conditions (*Clark* v. *Ryan* (1960) 103 CLR 486; *Weal* v. *Bottom* (1966) 40 ALJR 436). They were recognized as being the repositories of knowledge above and beyond that possessed by most lay persons. In the latter case, the Court determined that the evidence sought to be given about the tendency of semiarticulated vehicles to swing out when rounding a corner, that the tendency was more marked when the road surface was slippery and that therefore in such circumstances especial care needed to be shown, could be described as evidence of fact and therefore able to be given by persons not formally fulfilling the legal criteria for expertise.

Although experts are permitted to give evidence of opinion, they also give evidence in the form of fact. Such evidence can be particularly important – what the patient told a forensic physician, what a crime scene expert first saw upon arrival at a scene, what a forensic pathologist observed upon the body of a deceased. Frequently, it is the capacity of the trained 'scientist' to record observations meticulously, methodically and accurately that furnishes evidence critical to the resolution of key issues in both civil and criminal litigation.

See also: **Basic Principles of Forensic Science**. **Expert Witness**: Qualifications and Testimony. **Legal Aspects of Forensic Science**.

Further Reading

Chayko GM, Gulliver ED and Macdougall DV (1991) *Forensic Evidence in Canada*. Ontario: Canada Law Book Inc.

Coady CAJ (1992) *Testimony: A Philosophical Study*. Oxford: Clarendon Press.

Freckelton I (1987) *The Trial of the Expert*. Melbourne: Oxford University Press.

Freckelton I and Selby H (1999) *The Law of Expert Evidence*. Sydney: LBC Information Services.

Hodgkinson T (1990) *Expert Evidence: Law and Practice*. London: Sweet & Maxwell.

Imwinkelried EJ (1997) *The Methods of Attacking Scientific Evidence*, 3rd edn. Charlottesville: Michie.

Jasanoff S (1995) *Science at the Bar*. Cambridge, MA: Harvard University Press.

Jones CAG (1994) *Expert Witnesses*. Oxford: Clarendon Press.

Robertson B and Vignaux GA (1995) *Interpreting Evidence: Evaluating Forensic Science in the Courtroom*. Chichester: Wiley.

Saferstein R (1997) *Criminalistics: An Introduction to Forensic Science*. Upper Saddle River: Prentice Hall.

Stone J and Wells WAN (1991) *Evidence: Its History and Policies*. Sydney: Butterworths.

Tapper C (1995) *Cross on Evidence*, 8th edn. London: Butterworths.

Thayer JB (1898) *A Preliminary Treatise on Evidence at the Common Law*. New York: Augustus M Kelley, reprinted 1969.

Statistical Interpretation of Evidence/Bayesian Analysis

C G G Aitken, Department of Mathematics and Statistics, The University of Edinburgh, Edinburgh, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0503

Introduction

Bayesian analysis of the statistical interpretation of evidence relies on a rule relating the dependencies among uncertain events through conditional probabilities. This rule enables one to determine the value of evidence in the sense of the effect of the evidence on beliefs in an issue, such as the guilt or innocence of a defendant. The underlying ideas can be applied to categorical and continuous data and to personal beliefs. They can be used to ensure a logical structure to the evaluation of more than one item of evidence and to insure that the evidence is correctly interpreted.

Bayes' Rule

The Bayesian approach to the interpretation of evidence is named after the Reverend Thomas Bayes, a nonconformist preacher of the eighteenth century. He introduced an important rule, Bayes' rule, which showed how uncertainty about an event R can be changed by the knowledge of another event S:

$$P(R|S) = P(S|R)P(R)/P(S)$$

where P denotes probability and the bar | denotes conditioning. Thus $P(R \mid S)$ is the probability R occurs, given that S has occurred. Probabilities are values between 0 and 1; 0 corresponds to an event for which it is impossible that it will happen, 1 to an event which is certain to happen. An alternative version is the odds form, where $\bar{R}$ denotes the complement of R and $P(\bar{R}) = 1 - P(R)$. Then the odds in favor of R are $P(R)/P(\bar{R})$, denoted $O(R)$ and the odds in favor of $R \mid S$ are denoted $O(R \mid S)$. The odds form of Bayes' rule is then:

$$O(R|S) = \frac{P(S|R)}{P(S|\bar{R})} \times O(R).$$

In forensic science, R, S and $\bar{R}$ can be replaced in the odds form of Bayes' rule by E, H_p and H_d, where E is the scientific evidence, H_p is the hypothesis proposed by the prosecution and H_d is the hypothesis proposed by the defense.

$$O(H_p|E) = \frac{P(E|H_p)}{P(E|H_d)} \times O(H_p).$$

The left-hand-side of the equation is the odds in favor of the prosecution hypothesis after the scientific evidence has been presented. This is known as the posterior odds. The odds $O(H_p)$ are the prior odds (odds prior to the presentation of the evidence). The factor which converts prior odds to posterior odds is the fraction

$$\frac{P(E|H_p)}{P(E|H_d)}$$

known as the likelihood ratio or Bayes' factor. Denote this by V, for value. A value greater than one lends support to the prosecution hypothesis H_p and a value less than one lends support to the defense hypothesis H_d. Examples of H_p and H_d include guilt and innocence $(G, \bar{G})$, contact or no contact by the suspect with the crime scene $(C, \bar{C})$ or, in paternity

testing, the alleged father is, or is not, the true father (F, $\bar{F}$).

The scientific evidence is evaluated by determining a value for the Bayes' factor. It is the role of the forensic scientist to evaluate the Bayes' factor. It is the role of judge and jury to assess the prior and posterior odds. The scientist can assess how the prior odds are altered by the evidence but cannot assign a value to the prior or posterior odds. In order to assign such a value, all the other evidence in the case has to be considered.

If logarithms are used the relationship becomes additive:

$$log\{O(H_p|E)\} = \log\left\{\frac{P(E|H_p)}{P(E|H_d)}\right\} + \log\{O(H_p)\}.$$

This has the very pleasing intuitive interpretation of weighing evidence in the scales of justice and the logarithm of the Bayes' factor is known as the weight of evidence. Evidence in support of H_p (for which the Bayes' factor is greater than 1) will provide a positive value for the logarithm of the Bayes' factor and tilt the scales in a direction towards finding for H_p. Conversely, evidence in support of H_d (for which the Bayes' factor is less than 1) will provide a negative value for the logarithm of the Bayes' factor and tilt the scales in a direction towards finding for H_d. Evidence for which the value is 1 is neutral.

There is a difference between evaluation and interpretation. Evaluation is the determination of a value for the Bayes' factor. Interpretation is the description of the meaning of this value for the court. Both are considered here.

One of the advantages enjoyed by the Bayesian approach to evidence evaluation is the ability to combine so-called objective and subjective probabilities in one formula. Examples of this are given below in the section on transfer and background evidence. The objective probabilities are those obtained by analyses of data. For example, the relative frequencies of blood groups may be considered to give the probabilities that a particular individual, chosen at random from some relevant population (the one to which the criminal is thought to belong), is of a particular blood group. Subjective probabilities may be provided by the forensic scientist and can be thought of as a measure of the scientist's belief in the particular event under consideration. For example, when considering the transfer of evidence (such as blood or fibers), the scientist will have an opinion as to how likely (probable) it is that the evidence under examination would be transferred during the course of a crime and be recovered. This opinion can be expressed as a probability which represents the strength of the scientist's belief in the occurrence of the event about which he is testifying. The stronger the belief, the closer to 1 is the probability.

It is the subjective nature of the Bayesian approach which causes most controversy. Courts do not like subjectivity being expressed so explicitly (*R* v. *Adams*, 1996: 'The Bayes' theorem [rule] might be an appropriate and useful tool for statisticians but it was not appropriate for use in jury trials or as a means to assist the jury in their task.').

Another advantage of the Bayesian approach is the help it gives for the structuring of evidence. It insures logical consideration of the dependencies among different pieces of evidence. This is exemplified in the approaches based on influence diagrams or causal networks

Some terminology is required. Evidence whose origin is known is designated source evidence; evidence whose origin is unknown is designated receptor evidence. Other names have been suggested such as control, crime, known, questioned, recovered, suspect.

Early Pioneers of Bayesian Ideas

Examples of the early use of Bayesian reasoning in forensic science are found in relation to the Dreyfus' case. For example:

> An effect may be the product of either cause A or cause B. The effect has already been observed: one wants to know the probability that it is the result of cause A; this is the *a posteriori* probability. But I am not able to calculate this if an accepted convention does not permit me to calculate in advance the *a priori* probability for the cause producing the effect; I want to speak of the probability of the eventuality, for one who has never before observed the result. (Poincaré 1992)

Another quote, in the context of the Dreyfus' case is

> Since it is absolutely impossible for us [the experts] to know the *a priori* probability, we cannot say: this coincidence proves that the ratio of the forgery's probability to the inverse probability is a real value. We can only say: following the observation of this coincidence, this ratio becomes X times greater than before the observation. (Darboux et al. 1908)

Evaluation of Evidence

Categorical data

A fuller analysis of the Bayes' factor is given here to illustrate several underlying principles.

Let H_p be that the defendant was at the scene of the crime and that there was contact between the defendant and the scene resulting in the deposit of a bloodstain at the crime scene by the defendant; denote this hypothesis here by C. The defence hypothesis, H_d, is

that the defendant was not in contact with the crime scene and hence the bloodstain was deposited by someone else; denote this hypothesis by $\bar{C}$.

The evidence E has two parts:

- E_s: the blood group Γ of the defendant (source);
- E_c: the blood group Γ of the crime stain (receptor).

If the crime stain and the defendant had different blood groups then the blood group evidence would not be relevant. Let I denote the background information. This could include evidence concerning the ethnic group of the criminal, for example.

The value of the evidence is then

$$\frac{P(E|C,I)}{P(E|\bar{C},I)} = \frac{P(E_s,E_c|C,I)}{P(E_s,E_c|\bar{C},I)} = \frac{P(E_c|E_s,C,I)P(E_s|C,I)}{P(E_c|E_s,\bar{C},I)P(E_s|\bar{C},I)}.$$

Consider two assumptions:

1. The blood group of the defendant is independent of whether he was at the scene of the crime (C) or not ($\bar{C}$) and thus:
$$P(E_s|C,I) = P(E_s|\bar{C},I).$$
2. If the defendant was not at the scene of the crime ($\bar{C}$) then the evidence about the blood group of the stain at the crime scene (E_c) is independent of the evidence (E_s) about the blood group of the defendant and thus
$$P(E_c|E_s,\bar{C},I) = P(E_c|\bar{C},I).$$

Hence

$$V = \frac{P(E_c|E_s,C,I)}{P(E_c|\bar{C},I)}.$$

The above argument conditions on the blood group of the defendant. A similar argument, conditioning on the blood group of the crime stain shows that

$$V = \frac{P(E_s|E_c,C,I)}{P(E_s|\bar{C},I)}.$$

Let the frequency of Γ in the relevant population be γ. Assume the defendant is the criminal: the probability the crime stain is of group Γ, given the defendant is the criminal and is of group Γ is 1. Thus, the numerator of V is 1. If the defendant is not the criminal, then the probability the crime stain is of group Γ is just the frequency of Γ in the relevant population which is γ. Thus the denominator of V is γ. The value of the blood grouping evidence is then

$$V = \frac{1}{\gamma}.$$

As a numerical example, consider Γ to be the group AB in the OAB system with frequency $\gamma = 0.04$ in the relevant population. Then $V = 1/0.04 = 25$. The evidence is said to be 25 times more likely if the suspect were present at the crime scene than if he were not.

If the size of the population to which the criminal belongs (the relevant population) is $(N + 1)$, the prior odds in favor of the suspect equal $1/N$, assuming no evidence which favors any individual or group of individuals over any other, and the posterior odds in favor of the suspect equal $1/(N\gamma)$.

There is a more general result. Let $\pi_0 = P(C \mid I)$. For each of the other members of the population, let π_i be the probability that the ith member left the stain ($i = 1, \ldots, N$). Then

$$P(\overline{C}|I) = \sum_{i=1}^{N} \pi_i = 1 - \pi_0.$$

Let p_i denote the probability that the blood group of the ith person matches that of the crime stain: i.e. $P(E_c|E_s, \bar{C}, I) = p_i$, and let $w_i = \pi_i/\pi_0$. Then

$$P(C|E_s,E_c,I) = \frac{1}{1 + \sum_{i=1}^{N} w_i p_i}$$

and the posterior odds in favour of C are

$$\frac{P(C|E_s,E_c,I)}{P(\bar{C}|E_s,E_c,I)} = \frac{1}{\sum_{i=1}^{N} w_i p_i}.$$

If all of the w_i equal 1 and all of the p_i equal γ then the posterior odds are $1/(N\gamma)$ as before. The more general formulae for the posterior probability and odds can be used to account for evidence that the suspect has a relative, such as a brother, who has not been eliminated from the inquiry.

There are occasions when the numerator of the Bayes' factor is not 1. Consider an example where a crime has been committed by two people, each of whom has left a bloodstain at the scene. One stain is of group Γ_1, the other is of group Γ_2. A suspect is identified (for reasons unconnected with the blood evidence) and his blood is found to be of group Γ_1. The frequencies of these blood groups are γ_1 and γ_2, respectively. The two hypotheses to be considered are

- C: the crime stains came from the suspect and one other person;
- $\bar{C}$: the crime stains came from two other people.

The blood evidence E is

- the blood group (Γ_1) of the suspect, (E_s, source);
- the two crime stains of groups Γ_1 and Γ_2, (E_c, receptor).

Consider the numerator $P(E_c \mid E_s, C, I)$ of V. The stain of group Γ_1 came from the suspect since C is assumed true. Thus, the numerator is γ_2, the frequency of the other stain, the source of which is unknown. Now, consider the denominator $P(E_c \mid E_s, \bar{C}, I)$ of V. There are two criminals, of whom the suspect is not one. Denote these criminals A and B. Then either A is of group Γ_1 and B is of group Γ_2 or B is of group Γ_1 and A is of group Γ_2. These two events are mutually exclusive and both have probability $\gamma_1\gamma_2$. The probability of one event happening or the other, the denominator of V, is their sum, which is $2\gamma_1\gamma_2$. The ratio of the numerator to the denominator gives the value of V as

$$V = \gamma_2/(2\gamma_1\gamma_2) = 1/(2\gamma_2).$$

Compare this result with the single sample case. The value in the two-sample case is one half of that in the corresponding single sample case. This is reasonable. If there are two criminals and only one suspect, it is not to be expected that the evidence will be as valuable as in the case in which there is one criminal and one suspect. These ideas have been extended to cases in which there are general numbers of stains, groups and offenders.

Continuous data

A pioneering paper in 1977 by Dennis Lindley showed how the Bayes' factor could be used to evaluate evidence which was continuous data in the form of measurements. The measurements used by Lindley by way of illustration were those of the refractive index of glass. There were two sources of variation in such measurements, the variation within a window and the variation between different windows. Lindley showed how these two sources of variation could be accounted for in a single statistic. He was also able to account for the two factors which are of importance to a forensic scientist:

- the similarity between the source and receptor evidence;
- the typicality of any perceived similarity.

Normally, the prosecution will try to show that the source and receptor evidence have the same source, the defence will try to show they have different sources. Before Lindley's paper evidence had been evaluated in a two-stage process, comparison and typicality:

- are the source evidence and the receptor evidence similar – yes or no? (the comparison stage);
- if no, then assume they come from different sources; if yes, then determine the probability of similarity if the two pieces of evidence came from different sources (the assessment of typicality).

Interpretation is difficult. A small value for the probability of similarity if the two pieces of evidence come from different sources is taken to imply they come from the same source, which is fallacious reasoning, similar to that in the prosecutor's fallacy discussed below. Also, there is a cut-off point. Pieces of evidence whose measurements fall on one side of this point are deemed 'similar' and an assessment of the degree of similarity is made. Pieces whose measurements fall on the other side are deemed 'dissimilar' and it is assumed that they come from different sources. Lindley's approach achieves a continuous gradation from very similar to very dissimilar and encompasses these two stages into one formula. The formula has two components; one accounts for the comparison stage and one accounts for the assessment of typicality. Further developments of Lindley's approach have followed.

When the data are in the form of measurements the Bayes' factor is a ratio of probability density functions, rather than a ratio of probabilities.

Consider a set $\mathbf{x}$ of source measurements and another set $\mathbf{y}$ of receptor measurements of a particular characteristic, such as the refractive index of glass. For this example $\mathbf{x}$ would be a set of measurements of refractive indices on fragments of a broken window at the crime scene (source evidence) and $\mathbf{y}$ a set of measurements of refractive indices on fragments of glass found on a suspect (receptor evidence). If the suspect was at the crime scene then the fragments found on him could have come from the window at the crime scene, if he was not there then the fragments have come from some other, unknown, source. In general, the characteristic of interest may be parameterized, for example by the mean. Denote the parameter by θ. This parameter θ may vary from source (window) to source (another window). The evidence of the measurements is $E = (\mathbf{x}, \mathbf{y})$. The measurements $\mathbf{x}$ are from a distribution with parameter θ_1, say and the measurements $\mathbf{y}$ are from a distribution with parameter θ_2, say. If $\mathbf{x}$ and $\mathbf{y}$ come from the same source, then $\theta_1 = \theta_2$. In practice, the parameter θ is not known and the analysis is done with the marginal probability densities of $\mathbf{x}$ and $\mathbf{y}$. As before, consider two hypotheses

- C: the suspect was present at the crime scene;
- $\bar{C}$: the suspect was not present at the crime scene.

Continuous measurements are being considered so the probabilities P of the previous section are replaced by probability density functions f. Then, the value V of the evidence is given by

$$V = \frac{f(\mathbf{x}, \mathbf{y} | C, I)}{f(\mathbf{x}, \mathbf{y} | \bar{C}, I)},$$

which may be written as

$$V = \frac{\int f(\mathbf{y}|\theta)\mathbf{f}(\mathbf{x}|\theta)\mathbf{f}(\theta)d\theta}{\int f(\mathbf{x}|\theta)\mathbf{f}(\theta)d\theta \int \mathbf{f}(\mathbf{y}|\theta)\mathbf{f}(\theta)d\theta}.$$

Often, the distributions of ($\mathbf{x} \mid \theta$) and ($\mathbf{y} \mid \theta$) are assumed to be Normal, with θ representing the mean, varying from source to source, and the variance is assumed to be constant from source to source. Various possibilities have been assumed for the distribution of θ (with probability density function $f(\theta)$ which represents the variability among different sources. These include the Normal distribution or some estimation procedure such as kernel density estimation. Further developments are possible involving multivariate data.

Principles

Four principles arise from the application of these ideas.

First, population data need to exist in order to determine objective probabilities. For categorical data, such as blood groups, data are needed in order to know the frequency of a particular blood group. For measurement data, such as the refractive index of glass, data are needed in order to model the distribution of measurements, both within and between sources. This requirement for population data means that one needs to have a clear definition of the relevant population. The relevant population is determined by the hypothesis put forward by the defense concerning the circumstances of the crime and by the background information I. It is not necessarily defined by the circumstances of the suspect.

The second principle is that the distribution of the data has to be considered under two hypotheses, that of the prosecution and that of the defense.

The third principle is that evaluation is based on consideration of probabilities of the evidence, given a particular issue is assumed true.

The fourth principle is that the evaluation and interpretation of the evidence has to be conditional on the background information I.

Incorporation of Subjective Probabilities

Another advantage of the Bayesian approach is the ability to incorporate consideration of uncertain events, for which subjective probabilities of occurrence are required. Consider the assessment of the transfer and the presence of background material, innocently acquired.

A crime has been committed during which the blood of a victim has been shed. The victim's blood group (source evidence) is of group Γ. A suspect has been identified. A single bloodstain (receptor evidence) of group Γ is found on an item of the suspect's clothing. The suspect's blood group is not Γ. There are two possibilities: (A_0) the bloodstain on the suspect's clothing has come from some innocent source or (A_1) the bloodstain has been transferred during the commission of the crime. The hypotheses to consider are

- C: the suspect and victim were in contact;
- $\bar{C}$ the suspect and victim were not in contact.

Some additional probabilities are defined. These are $t_0 = P(A_0 \mid C)$ and $t_1 = P(A_1 \mid C)$, which denote the probabilities of no stain or one stain being transferred during the course of the contact. Also, let b_0 and b_1 denote the probabilities that a person from the relevant population will have zero bloodstains or one bloodstain on his clothing for innocent reasons. Let γ denote the frequency of blood group Γ in the relevant population. Then, it can be shown that

$$V = t_0 + \frac{t_1 b_0}{b_1 \gamma}.$$

The probabilities t_0, t_1, b_0 and b_1 are subjective probabilities, values for which are a matter of the forensic scientist's personal judgment. The probability γ is an objective probability, determined from examination of a sample from a relevant population and the determination from the sample of the relative frequency of group Γ in the population. The Bayesian approach enables these two different sorts of probabilities to be combined in a meaningful way. Without consideration of transfer and background probabilities, V would be equal to $1/\gamma$.

Combining Evidence

The representation of the value of the evidence as a Bayes' factor enables successive pieces of evidence to be evaluated sequentially. The posterior odds from one piece of evidence, E_1 say, become the prior odds for the next piece of evidence, E_2 say. Thus

$$\frac{P(G|E_1)}{P(\overline{G}|E_1)} = \frac{P(E_1|G)}{P(E_1|\overline{G})} \times \frac{P(G)}{P(\overline{G})}$$

and

$$\frac{P(G|E_1,E_2)}{P(\overline{G}|E_1,E_2)} = \frac{P(E_2|G,E_1)}{P(E_2|\overline{G},E_1)} \times \frac{P(G|E_1)}{P(\overline{G}|E_1)}$$
$$= \frac{P(E_2|G,E_1)}{P(E_2|\overline{G},E_1)} \times \frac{P(E_1|G)}{P(E_1|\overline{G})} \times \frac{P(G)}{P(\overline{G})}.$$

The possible dependence of E_2 on E_1 is recognized in the form of probability statements within the Bayes' factor. If the two pieces of evidence are independent

this leads to the Bayes' factors combining by simple multiplication:

$$\frac{P(E_1, E_2|G)}{P(E_1, E_2|\overline{G})} = \frac{P(E_1|G)}{P(E_1|\overline{G})} \times \frac{P(E_2|G)}{P(E_2|\overline{G})}.$$

If V_{12} is the Bayes' factor for the combination of evidence (E_1, E_2) and if V_1 and V_2 are the likelihood ratios for E_1 and E_2, respectively, then

$$V_{12} = V_1 \times V_2$$

If the weights of evidence (logarithms) are used different pieces of evidence may be combined by addition.

Bayesian analyses can also help to structure evidence when there are many pieces of evidence. This may be done using so-called influence diagrams or causal networks (**Fig. 1**). Influence diagrams represent the associations among characteristics (individual pieces of evidence) and suspects by a system of nodes and edges. Each characteristic or suspect is represented by a node. Causal connections are the edges of the diagram. Three stages may be considered. These are the qualitative representations of relationships as an influence diagram, quantitative expression of subjective beliefs as probabilities, and coherent evidence propagation through the diagram. Figure 1 represents the relationships in a hypothetical example where there are two suspects X and Y for the murder of a victim V and several pieces of evidence. There is eyewitness evidence E, fiber evidence F, and evidence that Y drives X's car regularly. T is a suggestion that Y picks up fibers from X's jacket.

Influence diagrams may be more of an investigatory tool than evidence as they can enable investigators to structure their investigation. A diagram can be constructed at the beginning of an investigation. Nodes can represent evidence that may or may not be found based on previous experience of similar crimes. The edges represent possible associations. Probabilities represent the levels of uncertainty attached to these associations. As evidence is discovered so the nodes and edges may be deleted or inserted. One would then hope that the belief in the guilt or innocence of suspects may become clearer as the probabilities associated with their nodes tend towards one or zero.

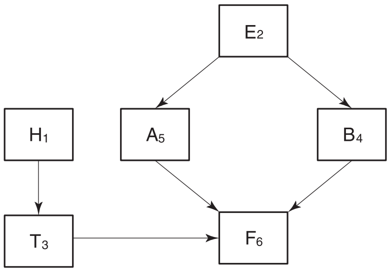


Figure 1 Causal network of nodes in the *fictitious example. A:* X committed the murder; B: Y committed the murder; E: Eyewitness evidence of a row between X, Y and the victim sometime before the commission of the crime; F: Fibers from a jacket similar to one found in the possession of X are found at the crime scene; H: Y drives X's car regularly; and T: Y picks up fibers from X's jacket. Reproduced from *Journal of the Forensic Science Society* (1989) p. 306.

Qualitative Scales

Qualitative scales are intended to make it easier to convey the meaning of the numerical value of the evidence. One such scale has four points:

- slight increase in support: $1 < V \leqslant 10^{\frac{1}{2}}$;
- increase in support: $10^{\frac{1}{2}} < V \leqslant 10^{3/2}$;
- great increase in support: $10^{3/2} < V \leqslant 10^{5/2}$;
- very great increase in support: $10^{5/2} < V$.

Such a scale is not sufficient to allow for the large values associated with DNA profiling, for example. A scale with more points on it and larger numerical values is more appropriate. This can be made manageable by the use of logarithms.

Interpretation

The Bayesian approach to the interpretation of evidence enables various errors and fallacies to be exposed. The most well-known of these are the prosecutor's and defender's fallacies. For example, a crime is committed. A bloodstain is found at the scene and it is established that it has come from the criminal. The stain is of a group which is present in only 1% of the population. It is also estimated that the size of the relevant population is 200 000. A suspect is identified by other means and his blood is found to be of the same group as that found at the crime scene.

The prosecutor argues that, since the blood group is present in only 1% of the population, there is only a 1% chance the suspect is innocent. There is a 99% chance he is guilty.

The defence attorney argues that, since 1% of 200 000 is 2000, the suspect is only one person in 2000. There is a probability of 1/2000 that he is guilty. Thus, the evidence of the blood group is not relevant to the case.

Consideration of the odds form of Bayes' rule explains these fallacies. Denote the blood type evidence by E. Let the two hypotheses be

- G: the suspect is guilty;
- $\bar{G}$: the suspect is innocent.

Then the odds form of Bayes' rule is that

$$\frac{P(G|E)}{P(\overline{G}|E)} = \frac{P(E|G)}{P(E|\overline{G})} \times \frac{P(G)}{P(\overline{G})}.$$

The Bayes' factor $P(E \mid G)/P(E \mid \bar{G}) = 1/0.01 = 100$. The posterior odds are larger than the prior odds by a factor of 100.

Consider the prosecutor's statement. He has said that the probability of guilt, after presentation of the evidence, is 0.99. In symbols, $Pr(G \mid E) = 0.99$ and, hence, $Pr(\bar{G} \mid E) = 0.01$. The posterior odds are 99, which is approximately 100. V is also 100. Thus, the prior odds are 1 and $Pr(G) = Pr(\bar{G}) = 0.5$. For the prosecutor's fallacy to be correct the prior belief is that the suspect is just as likely to be guilty as innocent. In a criminal court, this is not compatible with a belief in innocence until proven guilty.

The phrase 'innocent until proven guilty' does not mean that $P(G) = 0$. If that were the case, the odds form of Bayes' rule shows that, no matter how much evidence was led to support a hypothesis of guilt, $P(G \mid E)$ will remain equal to zero. Any person who believed that is unlikely to be permitted to sit on a jury.

The defense argues that the posterior probability of guilty $P(G \mid E)$ equals 1/2000 and, hence, $P(\bar{G} \mid E)$ equals 1999/2000. The posterior odds are 1/1999, which is approximately 1/2000. Since the posterior odds are bigger by a factor of 100 than the prior odds, the prior odds are 1/200 000, or the reciprocal of the population size. The defense is arguing that the prior belief in guilt is approximately 1/200 000. This could be expressed as a belief that the suspect is just as likely to be guilty as anyone else in the relevant population. This seems a perfectly reasonable definition of 'innocent until proven guilty'. The fallacy arises because the defense then argues that the evidence is not relevant. However, before the evidence was led, the suspect was one of 200 000 people, after the evidence was led he is only one of 2000 people. Evidence which reduces the size of the pool of potential criminals by a factor of 100 is surely relevant.

Other errors have been identified. The ultimate issue error is another name for the prosecutor's fallacy. It confuses the probability of the evidence if a defendant is innocent with the probability he is innocent, given the evidence. The ultimate issue is the issue proposed by the prosecution of which it is asking the court to find in favor. The source probability error is to claim the defendant is the source of the evidence. This would place the defendant at the scene of the crime but would not, in itself, be enough to show that he was guilty. The probability (another match) error assigns the relative frequency of a characteristic to the probability that another person has this characteristic. The numerical conversion error equates the reciprocal of the relative frequency to the number of people that have to be examined before another person with the same characteristic is found.

High values for the evidence provide strong support for the prosecution evidence. They are not sufficient in themselves to declare a defendant guilty. The prior odds have to be considered. Very high values for the evidence, when combined with very small values for prior odds, may produce small values for the posterior odds.

See also: **Deoxyribonucleic Acid**: Statistical Analysis.

Further Reading

Aitken CGG (1995) *Statistics and the Evaluation of Evidence for Forensic Scientists*. Chichester: Wiley.

Aitken CGG and Stoney DA (eds) (1991) *The Use of Statistics in Forensic Science*. Chichester: Ellis Horwood.

Aitken CGG and Gammerman A (1989) Probabilistic reasoning in evidential assessment. *Journal of the Forensic Science Society* **29**:303–316.

Darboux, Poincaré (1908) Examen critique des divers systèmes ou études graphologiques auxquels a donné lieu le bordereau. In: *Affaire Dreyfus – la révision du procès de Rennes – enquête de la chambre de la Cour de cassation*. Paris: Ligue française des droits de l'homme et du citoyen.

Dawid AP and Evett IW (1997) Using a graphical method to assist the evaluation of complicated pieces of evidence. *Journal of Forensic Sciences* **42**:226–231.

Evett IW (1984) A quantitative theory for interpreting transfer evidence in criminal cases. *Applied Statistics* **33**:25–32.

Evett IW (1987) On meaningful questions: a two-trace transfer problem. *Journal of the Forensic Science Society* **27**:375–381.

Evett IW (1987) Bayesian inference and forensic science: problems and perspectives. *The Statistician* **36**:99–105.

Evett IW and Weir BS (1998) *Interpreting DNA Evidence*. Sunderland, MA: Sinauer.

Good IJ (1991) Weight of evidence and the Bayesian likelihood ratio. In: Aitken CGG and Storey DA (eds) *The Use of Statistics in Forensic Science*. Chichester: Ellis Horwood.

Lindley DV (1977) A problem in forensic science. *Biometrika* **64**:207–213.

Poincaré H (1992) *La Science et l'hypothèse*. Rueil-Malmaison: Editions de la Bohème.

Robertson BW and Vignaux GA (1995) *Interpreting Evidence: Evaluating Forensic Science in the Courtroom*. Chichester: Wiley.

Schum DA (1994) *The Evidential Foundations of Probabilistic Reasoning*. New York: Wiley.

Taroni F and Aitken CGG (1998) Probalistic reasoning in the law. Part 1: Assessment of probabilities and explanation of the value of DNA evidence. *Science and Justice* **38**:165–177.

Taroni F and Aitken CGG (1998) Probabilistic reasoning in the law. Part 2: Assessment of probabilities and explanation of the value of trace evidence other than DNA. *Science and Justice* **38**:179–188.

Taroni F, Champod C and Margot P (1998) Forerunners of Bayesianism in early forensic science. *Jurimetrics* **38**:183–200.

Thompson WC and Schumann EL (1987) Interpretation of statistical evidence in criminal trials. The prosecutor's fallacy and the defence attorney's fallacy. *Law and Human Behaviour* **11**:167–187.

EXPERT WITNESS

Qualifications and Testimony

C Henderson, J. D. Nova Southeastern University Shepard Broad Law Center, Fort Lauderdale, FL, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0504

Introduction

Expert testimony continues to play an important role as we enter the twenty-first century; however, an expert's scientific findings, no matter how important, have no real meaning in law until they are presented to the judge or jury.

This article will first examine the admissibility of scientific evidence, including the most recent United States Supreme Court decisions and proposed changes to the Federal Rules of Evidence. The article will next address experts' qualifications and how to project credibility and expertise, through those qualifications, to a judge or jury. Finally the article will examine how jurors perceive expert witnesses. Various studies, conducted before and after the infamous O.J. Simpson trial, will be discussed in order to examine whether jurors' perceptions of experts have changed.

Admissibility of Expert Testimony

The admissibility of expert testimony is an issue to be decided by the judge presiding over the case. In the United States there are two tests for admissibility, the older *Frye* test, or 'general acceptance' test, which received its name from the 1923 case of *Frye* v. *United States*, 293 F. 1013 (D.C. Cir. 1923). *Frye* dealt with the admissibility of the precursor of the polygraph. The Court stated the test as follows:

> Just when a scientific principle or discovery crosses the line between the experimental and demonstrable stages is difficult to define. Somewhere in this twilight zone the evidential force of the principle must be recognized, and while courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs.

Id. at 1014. The Court held that the systolic blood pressure test had not yet gained recognition in the physiological and psychological communities, therefore the evidence was not admitted.

The *Frye* test was the test used by the majority of states and the federal courts in the United States for many years. In 1975, the Federal Rules of Evidence were enacted. Rule 702 provides:

> If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise.

Therefore, to be admissible, such evidence need be helpful, relevant and reliable.

After 1975, there arose a question whether Rule 702 superseded the *Frye* test. In 1993, the United

States Supreme Court answered the question in the affirmative in *Daubert* v. *Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579, 113 S.Ct. 2786 (1993). In *Daubert*, the plaintiffs alleged that the ingestion of Benedictin, a prescription antinausea drug marketed by Merrell Dow Pharmaceuticals, caused birth defects such as limb deformation. The Supreme Court held that proof that establishes the scientific reliability of expert testimony must be produced before it may be admitted. *Daubert* requires that the judges become 'gatekeepers' of scientific evidence. *Daubert* requires the trial judges to decide whether the expert testimony will assist the trier of fact and whether it amounts to scientific knowledge, i.e. is the theory or test derived by the scientific method? The United States Supreme Court listed factors that the judges should consider in making their decision; however, the Court stated the list is not a definitive checklist, nor is one factor determinative:

1. Whether the proposition is testable.
2. Whether the proposition has been tested.
3. Whether the proposition has been subjected to peer review and publication.
4. Whether the methodology or technique has a known or potential error rate.
5. Whether there are standards for using the technique.
6. Whether the methodology is generally accepted.

When the *Daubert* case was remanded to the Ninth Circuit Court of Appeals to apply the factors, the Court held that the evidence did not amount to scientific knowledge and was therefore inadmissible. The Ninth Circuit also noted that the technique was developed solely for the litigation in the case and the proffered expert testimony was not based on independent research, thus not subjected to peer review and publication.

The *Daubert* ruling had a great impact on states that have evidence codes modeled after the Federal Rules of Evidence (by mid-1993 35 states had such evidence codes). At present, the majority of states have adopted the *Daubert* test of admissibility of scientific evidence, while a minority has retained the *Frye* test.

In 1999 the United States Supreme Court had to decide whether the 'gatekeeper' function of courts, as outlined in *Daubert*, reached the testimony of all experts, not just 'scientific' experts. *Kumho Tire Co. Ltd.* v. *Carmichael*, 119 S.Ct. 1167 (1999), involved a car accident that occurred after a tire ruptured. The survivors and decedent's representative sued the tire's manufacturer and distributor, alleging a defect in the tire caused the rupture, which led to the accident. The trial court excluded the plaintiffs' expert's testimony for failing to meet the *Daubert* criteria. The Eleventh Circuit Court of Appeals reversed, reviewing *de novo* the district court's decision to apply *Daubert*. The Circuit Court ruled that the *Daubert* test is restricted to those cases involving 'scientific' evidence and not those involving skill or experience-based knowledge.

The Supreme Court reversed and held that the admissibility of expert testimony based upon technical or specialized knowledge should be subjected to the analysis and facts set forth in *Daubert* v. *Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993). The Supreme Court also reaffirmed that the *Daubert* factors were not restrictive, but flexible and should be applied on a case-by-case basis.

Commentators have agreed that the greatest impact of the *Kumho Tire* decision will be in the areas of 'soft science': handwriting, fingerprints, psychological testing, medical and psychiatric testimony, arson investigation, accident reconstruction, bitemarks and drug recognition experts. See 'Court Developments: Daubert Applies to *all* Experts, Not Just "Scientific" Ones, High Court Holds,' 13, No. 7 The Criminal Practice Report 129,132 (April 7, 1999). Thus, the *Kumho Tire* decision is likely to open the door to litigation in these areas. Additionally, this decision will foster changes in the admissibility of, at least, some expert evidence.

At the time of the writing of this article there was a proposed amendment to Rule 702 of the Federal Rules of Evidence which reads as follows:

> If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise, *provided that (1) the testimony is sufficiently based upon reliable facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case.*

This proposed amendment was drafted in response to the *Daubert* decision and to the many cases applying *Daubert*. 'The amendment does not distinguish between scientific and other forms of expert testimony,' Proposed Advisory Committee Note. The Committee Note also states that the court's 'gatekeeper' function applies to testimony by any expert and that '[t]he trial judge in all cases of expert testimony must find it is properly grounded, well-reasoned, and not speculative before it can be admitted.'

The Committee Note points out that if a witness is relying primarily on experience to reach an opinion, the witness must explain how that experience leads to the opinion. The Committee Note cautions that the more controversial and subjective an opinion, the

more likely the testimony should be excluded as unreliable. The amendment requires, in its revised format, that expert testimony must be based upon reliable and sufficient underlying 'facts or data'. This 'data' includes reliable opinions of other experts.

Once a judge determines that scientific evidence is admissible, the judge must still address whether the probative value of the evidence is outweighed by prejudice or whether such evidence may confuse the jury. Federal Rule of Evidence 403 provides:

> Although relevant, evidence may be excluded if its probative value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury, or by considerations of undue delay, waste of time, or needless presentation of cumulative evidence.

The admissibility of expert evidence is a question of law that the judge decides. A judge's decision will only be overturned on appeal if the appellate court determines the judge abused his or her discretion. The United States Supreme Court adopted the abuse of discretion standard for reviewing a trial court's admissibility decision under *Daubert* in *General Electric Company* v. *Joiner*, 118 S.Ct. 512 (1997).

Qualifications

Once the issue of admissibility of scientific evidence is resolved, the judge must determine whether a witness is qualified to render an opinion as an expert. In the United States, the Federal Rules of Evidence state that a witness may qualify as an expert on the basis of knowledge, skill, training, experience or education.

An expert witness need only possess one of these traits for the judge to find the expert qualified to give an opinion. In making this evaluation, the judge may consider the expert's educational background, work experience, publications, awards, teaching or training positions, licenses or certification, speaking or other professional engagements, prior expert witness testimony, and membership of professional associations. In the majority of jurisdictions, the determination of a witness's qualification to express an expert opinion is within the discretion of the trial judge and will only be overturned for abuse of discretion. In the minority of jurisdictions, the appellate courts will not reverse the judge's decision in the absence of a clear showing of error. Often, the expert may have to educate the attorney proffering the expert regarding the significance of particular experience, achievements and certifications to ensure that they receive the appropriate emphasis. An expert must be prepared to explain board certification and licensure in detail, including the recertification and relicensing requirements.

Experience as an Expert Witness

Experience and training are often more significant than academic background and are accorded more weight by jurors, according to a jury study evaluating fingerprint experts. See Charles Illsley, *Juries, Fingerprints and the Expert Fingerprint Witness*, US Department of Justice (1987). However, experience as an expert witness, standing alone, does not qualify someone as an expert in later cases. For example in *Bogosian* v. *Mercedes-Benz of N. Am., Inc.*, 104 F.3d 472, 477 (1st Cir. 1997), the court rejected an opinion of a witness who had testified as an expert 126 times. One court even noted 'it would be absurd to conclude that one can become an expert by accumulating experience in testifying' – *Thomas J. Kline, Inc.* v. *Lenillard, Inc.*, 878 F.2d 791, 800 (4th Cir. 1989). Conversely, a lack of prior experience as an expert witness does not disqualify one from testifying as an expert. *United States* v. *Locascio*, 6 F.3d 924, 937 (2nd Cir. 1993) ('... even the most qualified expert must have his first day in court.').

Education/Training

An expert may be qualified based on his or her academic credentials, i.e. undergraduate, graduate and/or postgraduate work. While distance learning is the way of the future, the proliferation of the Internet has rekindled the old-fashioned diploma mill, so one needs to be careful to participate only in accredited programs. An expert also needs to keep up-to-date with developments in his or her field by reading the literature (journals, treatises, newsletters and books), as well as through continuing education, joining professional societies and attending professional meetings.

Teaching experience is another of the qualifications that judges will evaluate; not only teaching as a regular faculty member, but guest lecturing, visiting professorship and teaching of continuing education and short courses all weigh in as credentials.

An expert also needs to keep researching and publishing. In the past, one of the criticisms of the document examination field was a lack of research and studies. *See* D. Michael Risinger, et al, "Exorcism of Ignorance as a Proxy for Rational Knowledge: The Lessons of Handwriting Identification Expertise,"' 137 U. PA.L.REV. 731 (1989) and D. Michael Risinger and Michael J. Saks, "Science and Nonscience in the Courts: Daubert Meets Handwriting Identification Expertise," 82 IOWA L. REV. 21 (1996). That criticism has been rebutted. *See* Andre A. Moenssens, "Handwriting Identification in a Post-*Daubert* World," 66 U.M.K.C. L.REV. 251 (1997)

but the debate is still on-going. For a response to Professor Moenssens' article, see D. Michael Risinger, Mark Denbenbeaux, Michael J. Saks in "Brave New 'Post-Daubert World,' – A Reply to Professor Moenssens," 29 Stan L. Rev. 405 (1998).

Membership in Professional Associations

A study published by the *Juries, Fingerprints and the Expert Fingerprint Witness* United States Department of Justice (1987) found that jurors perceived those fingerprint experts who belonged to professional associations as more credible than other experts, and presumed experts would belong to such groups. Joining such associations aids in making contacts, interaction with peers and provides speaking and publishing opportunities as well as educational opportunities. At professional meetings experts have opportunities to hear and discuss cutting edge ideas and theories with their peers.

Annual payment of dues alone, in order to be a member of a professional association, is not as prestigious as having to be invited for membership or needing referees or requiring an examination for membership. It is important to remain active and participate aggressively in professional societies. The expert's credibility is diminished if the expert has not attended a professional meeting recently.

It is wise for an expert to be selective about which professional associations to join. The February 8, 1999 issue of the *Wall Street Journal* of 8 February 1999 noted that the American College of Forensic Examiners, a 6-year-old organization, which already has 12 000 members, has 'mail-order' credentialing, for which the applicants must only pay $350.00 and pass an ethics exam. There were no examinations for board certifications in various specialties during their 'waiver of examination' periods – all one needed to do was verify possession of 200 'experience points' and send in a fee.

The founder of the group is currently seeking to establish an Internet-based educational program that would offer PhDs in forensic science. So far, no college or university has decided to host the program. Elizabeth MacDonald, 'The Making of and Expert Witness: It's In the Credentials,' WALL ST J. Feb. 8, 1999, at B1.

Increased Scrutiny of Experts

Experts have come under increased scrutiny for either fabricating or inflating their qualifications. For example, in 1998, in Florida, a person who had been testifying as an expert in toxicology for 3 years, for both the prosecution and defense in criminal cases, was prosecuted for perjury for testifying with fraudulent credentials. A prosecutor noticed some discrepancies between two of the expert's resumés and began checking into his claimed credentials. The expert claimed to possess a masters and doctorate degree from Florida Atlantic University. The registrar's office had no record of his attending or of receiving a degree from the university. In fact, the university does not even offer a PhD. in organic chemistry, the program from which the expert claimed to graduate. The expert also handed out a copy of his master's degree dated 1971. It was signed by 'Lawton Chiles, Governor'. Chiles was a United States Senator in 1971. He did not serve as Florida's Governor until 1991. The expert was eventually charged with three counts of second degree perjury (one count for each first-degree murder case in which the expert testified as a defense witness). The 'expert' pled guilty and received a sentence of 3 years in prison, followed by 5 years probation. *See* Henry Fitzgerald, Jr., "Phony 'expert' jailed for 3 years," FT. LAUDERDALE SUN SENTINEL, Dec. 1, 1998, at 3D. In addition to perjury prosecutions for false qualifications, some jurisdictions will also prosecute for academic fraud. For example, in Florida, a person making a false claim, either orally or in writing, to possession of an academic degree, or title associated with such a degree (e.g. PhD), is guilty of a first-degree misdemeanor. In addition, a licensed person may have his or her license to practice suspended or revoked. *See* Fla. Stat. §817.567 (1999).

In another matter, a Harvard Medical Professor was sued for trademark infringement for falsely claiming to be board certified by the American Board of Psychiatry and Neurology (ABPN) in five trials. The Board sought to seize the expert witness fees and treble damages – *ABPN* v. *Johnson-Powell*, 129 F.3d 1 (1997). (While seemingly in agreement that the expert had committed infringements, the trial court denied relief because it believed she was unlikely to infringe in the future. The appellate court affirmed, stating that the trial court did not abuse its discretion.)

Courts have also overturned convictions where the experts testified outside their field of expertise. For example, in *Gilliam* v. *State*, 514 So.2d 1098 (Fla. 1987), the court held the medical examiner was not qualified as an expert in shoe pattern analysis; therefore, it was an error for the trial court to allow her to testify that the defendant's sneaker left marks on the decedent's body. See also, *Kelvin* v. *State*, 610 So.2d 1359 (Fla. App. 1992), in which an evidence technician was found not qualified to give expert testimony about the trajectory of bullets depicted by dowels

stuck into bullet holes in a sofa at a crime scene, as he was not a crime scene reconstructionist and had no training in ballistics.

Weight of the evidence

Once a judge decides that an expert may testify, the jury must then decide the weight to accord the expert's opinion. Expert witnesses and attorneys should be aware of the studies that have been conducted regarding jurors' perceptions of expert witnesses. These studies have shown that jurors give great weight to expert testimony. Many of these studies discuss jurors' evaluations of the experts' qualifications as well as the experts' appearance, demeanor and communication skills.

In 1978 Rich Tanton published a study in the Journal of Forensic Sciences, 'Jury Preconception and Their Effect on Expert Scientific Testimony,' 24 J. FOR SCIS. 681 (1978). Tanton found that jurors held stereotypical views of experts: for example, the male expert was expected to be 44 years old, upper middle class, white, neat, intelligent, wearing a dark suit and glasses, professional, calm and serious; the female expert was expected to be white, 37 years old, dressed conservatively, neat, pleasant and honest.

In 1984, Saks and Wissler conducted a telephone survey of potential jurors. Michael Saks and Roselle Wissler, 'Legal and Psychological Bases of Expert Testimony: Surveys of the Law and Jurors,' 2 BEHAV. SCI. & L. 435 (1984). The respondents were asked the following: whether they would believe testimony from the experts; whether they perceive the testimony to be honest; and whether they perceive the witnesses from different fields to have adequate experience to testify. The study concluded that physicians, chemists and firearm experts ranked the highest in believability, honesty and experience. Next highest ranked were accountants, psychiatrists, psychologists and eyewitnesses. Lowest ranked were police officers, handwriting experts and polygraph experts.

Dr Joseph Peterson conducted three studies on forensic evidence and the courts from 1984 to 1987. Joseph Peterson, S. Mihajlovie, M. Gilliland, *Forensic Evidence and the Police: The Effects of Scientific Evidence on Criminal Investigations*, National Institute of Justice Research Report (Oct. 1984); Joseph Peterson, J. Ryan, P. Houlden, S. Mihajlovic, *Forensic Science and The Courts: The Uses and Effects of Scientific Evidence in Criminal Case Processing*, Chicago Center for Research in Law and Justice, U. 111. At Chicago (1986); Joseph Peterson, *Use of Forensic Evidence by the Police and Courts*, National Institute of Justice Research in Brief (Oct. 1987) (Ser. No.: NCJ 107206). His studies concluded jurors accorded great weight to expert testimony; the author of these found that in criminal cases in which expert witnesses testified for the prosecution, jurors were more likely to render guilty verdicts.

In 1992, the *National Law Journal*, a legal publication in the United States, conducted a survey of jurors and their views of the jury system. Joan Cheever and Joanne Naiman, 'The View from the Jury Box,' Vol. 15, No. 25 NAT'L. L.J., Feb. 22, 1993, at s2 col.1. The study concluded that jurors were influenced by expert witnesses and accorded their opinions great weight. In all the civil and criminal cases surveyed, 89% of the jurors thought the experts were believable. Overall 71% of the jurors said the experts made a difference to the verdict.

A more recent study of experts and jurors' preconceived notions of experts was carried out in 1994 by Daniel W. Shuman, Elizabeth Whitaker and Anthony Champagne, 'An Empirical Examination of the Use of Expert Witnesses in the Courts – Part II: A Three City Study,' 34 JURIMETRICS J. 193 (1994). They conducted an extensive survey of American lawyers, judges, jurors and expert witnesses. A fascinating aspect of the study was the tabulation of the characteristics of experts that jurors considered important in determining the experts' credibility. They determined the willingness to draw firm conclusions and the ability to convey technical information nontechnically were the most important characteristics of a credible expert.

Shuman, Champagne and Whitaker published another study, 'Assessing the Believability of Expert Witnesses: Science in the Jurybox,' 37 JURIMETRICS J. 23 (1996), which concluded that an expert witness's believability was linked to the expert's qualifications, familiarity with the facts of the case, good reasoning and perceived impartiality. Jurors were also influenced by independent research that corresponded with the expert's opinion.

A 1997 study of jurors' perceptions of expert witnesses in death penalty cases by Scott F. Sundby, 'The Jury as Critic: An Empirical Look at How Capital Juries Perceive Expert and Lay Testimony,' 83 VA. L. REV. 1109 (1997) found that defenses based solely on expert testimony are likely to fail, but defenses that integrate expert testimony with persuasive lay testimony are more likely to prevail.

A 1998 *National Law Journal*/Decision Quest study Peter Aronson, David E. Rovella and Bob Van Voris, 'Jurors: A Biased, Independent Lot,' Nat'1 L.J., Nov. 2, 1998, at A1, exposes jurors as a more skeptical, cynical group. Among the findings, the study concluded that 50% of those surveyed think expert witnesses say only what they are paid to say; 33% do not believe police testimony; and 75% said they

would set aside what a judge says the law requires and reach a verdict the jurors feel is right.

Conclusion

Expert testimony will continue to play an important role in the future. To be a more effective witness, an expert should be aware of the legal tests for admissibility of scientific evidence, the factors courts will evaluate in order to determine whether an expert is qualified to testify and jurors' perceptions of experts.

See also: **Legal Aspects of Forensic Science**.

Further Reading

Feder HA (1993) *Succeeding as an Expert Witness*. Colorado Springs: Tageh Press.

Froede RC (1997) *The Scientific Expert in Court*. Washington DC: AACC Press.

Lubet S (1998) *Expert Testimony, a Guide for Expert Witnesses and the Lawyers who Examine Them*. Notre Dame: NITA.

Moenssens A, Starrs J, Henderson C and Inbau F (1995) *Scientific Evidence in Civil and Criminal Cases*, 4th edn. New York: Foundation Press.

Rossi F (1991) *Expert Witnesses*. Chicago: American Bar Association.

EXPLOSIVES

Contents

Analysis
Bomb-scene Management
Commercial
Mechanism of Explosion
Military

Analysis

T Tamiri, Israel Police, Division of Identification and Forensic Science, Jerusalem, Israel

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0505

Caution!

Some explosives are extremely sensitive to sparks, heat, shock or friction; necessary precautions should therefore be taken by the analyst while handling explosives.

Introduction

General

An explosive is defined as a substance or mixture of substances which may be made to undergo a rapid chemical change without an outside supply of oxygen, with the liberation of large quantities of energy, generally accompanied by the evolution of hot gases. An explosive mixture must contain an oxidizing agent and a reducing agent.

Explosives are used legally in industry and in the military but also illegally in terrorist and criminal activity. Some explosive compounds have other uses such as in medicine (some nitrate esters) or in the paint industry (cellulose nitrate).

Military and industrial explosives are routinely analyzed after their production for quality control. In this area, apart from qualitative analysis, the analysis usually includes quantitation of explosives, by-products and impurities in order to meet accepted specifications. Other areas which deal with the analysis of explosives are forensic science, environmental analysis and in certain cases, toxicology.

Forensic analysis of explosives deals with the identification of unreacted explosives as well as with postexplosion identification. The identification of an unexploded explosive is carried out to prove its possession or its intended use.

In postexplosion analysis the situation is different; when an explosion has already occurred, it is

reasonable that an explosive was involved. It may seem unnecessary to perform postexplosion analysis, but such analyses have the highest priority in most forensic laboratories. The reason is that information about the explosives involved can be of great assistance to the investigation. Sometimes it is not unequivocally clear whether the explosion was initiated by a high explosive or by the ignition of a fuel–air mixture ('vapor explosion'). When an explosive is identified in residues it may strongly suggest that it caused the explosion. On the other hand if no explosive is identified it may suggest that no explosive was involved (e.g. 'vapor explosion') but it may also be that the analysis was unsuccessful. Sometimes the results of the analysis can direct the investigator as to whether the explosion was carried out by terrorists or by criminals unrelated to terrorist activity. Certain types of explosives have been typical to terrorist groups (e.g. 'Semtex', especially before the fall of the 'iron curtain'). In rare cases the type of explosive may even hint to a certain organization. Another reason to pursue postexplosion analysis is the need for law-enforcement agencies to know what materials are used by criminals or terrorists. This information may help in connecting between different cases and also to realize that some materials are not as 'innocent' as they seem, but are starting materials for the preparation of these explosives (e.g. acetone and hydrogen peroxide for the preparation of triacetonetriperoxide (TATP)).

Another very important type or work is the trace analysis of explosives on suspects' hands, and on items and premises which may be related to suspects. Although not a postexplosion situation, the procedures used in such analyses are similar to those used in postexplosion analysis. The analysis in these cases is difficult because it usually deals with trace amounts of unreacted explosive mixed with large amounts of contaminants. The ultimate goal of a forensic analyst is to provide an expert opinion for a court of law. Wrong results may lead to a gross injustice, where innocent people can be found guilty. This dictates the need to adhere to extremely strict criteria for the safe identification of explosives.

Classification of explosives

Many compounds, which may be classified as explosives, are listed in the literature. In practice, the forensic analyst routinely encounters only few compounds. Explosives may be classified according to their chemical structure, use, their place in the detonation chain and their explosive properties. These widely used classifications are given, along with a few examples:

Chemical structure

- Organic nitro explosives:
  Nitroaromatic: 2,4-dinitrotoluene (2,4-DNT), 2,4,6-trinitrotoluene (TNT), picric acid;
  Nitrate esters: ethyleneglycol dinitrate (EGDN), glycerol trinitrate (NG), pentaerythrithol tetranitrate (PETN), cellulose nitrate (NC);
  Nitramines: 1,3,5-trinitro-1,3,5-triazacyclohexane (RDX), 1,3,5,7-tetranitro-1,3,5,7-tetrazacyclooctane (HMX);
- Organic peroxides: 3,3,6,6,9,9-hexamethyl-1,2,4,5,7,8-hexaoxacyclononane (TATP), 3,4,8,9,12,13-hexaoxa-1,6-diazabicyclo [4,4,4] tetradecane (HMTD);
- Inorganic salts: ammonium nitrate (AN);
- Mixtures of oxidizing and reducing agents: black powder (potassium nitrate, sulfur and charcoal), potassium chlorate and sugar.

Structural formulae of organic explosives representing each chemical group are shown in **Fig. 1**.

Use

- Military explosives: TNT, PETN, RDX
- Industrial or commercial explosives: dynamites, AN, emulsion explosives.
- Illegally manufactured improvised explosives: TATP, potassium chlorate and sugar mixture.

Place in detonation chain

- Primary explosives: mercury fulminate, lead styphnate, lead azide, dinol.
- Boosters: PETN.
- Main charge: TNT, RDX.

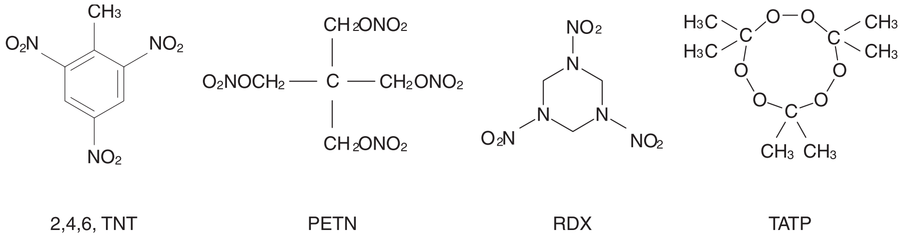


Figure 1 Structural formulae of some organic explosives.

Explosive properties

- High explosives: PETN, RDX
- Low explosives (propellants): smokeless powder, black powder.

Explosive formulations usually contain other compounds such as sensitizers, desensitizers, stabilizers (e.g. diphenylamine (DPA), ethyl centralite (EC)), plasticizers (e.g. phthalate and sebacate esters) and other additives.

Methods and Procedures

Analytical methods

Various analytical methods have been used for the analysis of explosives: chemical tests (based on color reaction); chromatographic methods such as thin layer chromatography (TLC), column chromatography, gas chromatography (GC), high pressure liquid chromatography (HPLC), capillary electrophoresis (CE) and ion chromatography (IC); and spectral methods such as infrared (IR), nuclear magnetic resonance (NMR), mass spectrometry (MS), scanning electron microscopy–energy dispersive X-ray spectroscopy (SEM/EDX) and X-ray diffraction (XRD). Hyphenated methods, usually on-line combinations of chromatographic and spectroscopic methods (e.g. GC/MS, HPLC/MS) are powerful analytical techniques which have become widely used in forensic laboratories.

Analytical procedures

The analytical procedures used in forensic laboratories for the analysis of explosives are based on the methods described above. Naturally, a procedure for postexplosion analysis is different from the analysis of intact explosives.

In the latter case the analysis is based on a normal methodology for unknown samples. The procedure may include spot tests, chromatographic methods and spectrometric methods. It is essential for the forensic laboratory to have a library of spectral data of all common explosives and related compounds and preferably also authentic samples of the explosives. The analysis usually includes identification of the sample but sometimes quantitation is also required (e.g. when a common origin of samples is suspected).

Procedures for postexplosion analysis are much more complicated, including recovery techniques as well as methods of identification. Recovery methods may include microscopic examination, headspace sampling, adsorption and extraction procedures. The identification, as in many areas in forensic analysis, may start with a screening method followed by a suitable confirmation method. Postexplosion analysis, including a flow chart, is discussed separately later.

The methods in a procedure may also be derived from the properties of the explosives. Different explosives have different chemical and physical properties. Some explosives are highly volatile (e.g. EGDN), some undergo easy sublimation (e.g. TATP) and some are nonvolatile (e.g. HMX). Some explosives are thermally labile and decompose when heated. Knowing the properties of the different explosives is relevant to the choice of the most suitable method for the recovery and the analysis of specific explosives. For example, analytical methods which involve evaporation of compounds (e.g. GC or GC/MS) may be unsuitable for nonvolatile explosives.

Finally and not less important, the procedures used in each laboratory are influenced by the financial ability to purchase sophisticated instrumentation and by the professional skill of the staff.

Methods, advantages and limitations

Chemical tests Chemical tests, also referred to as spot tests or color tests, take advantage of a color produced by a reaction between a reagent and an analyte. Some well-known color reactions (sometimes in a modified version) are used in the analysis of explosives. These color reactions are widely used by many forensic laboratories as presumptive tests, for screening and for field tests.

Some widely used spot tests are described according to the classes of explosives. Di- and trinitroaromatic compounds develop colors with some basic solutions: TNT develops a purple–brown color whereas 2,4-DNT and 2,6-DNT develop a yellowish color when reacted with KOH in ethanol (e.g. 3%). The colors developed in the reaction between polynitroaromatic compounds and bases are sometimes attributed to the formation of the so called 'Meisenheimer complexes'.

The Griess reaction is a well-established, highly specific color reaction for the identification of nitrite ions. In this reaction, nitrite ion reacts with an aromatic amine, such as sulfanilamide, in an acidic medium to form a diazonium ion. This ion is then coupled with a suitable active aromatic compound, such as *N*-1-naphthylethylenediamine, to produce an azo compound, which has a characteristic purple color. Nitrate esters and nitramines produce NO_2^- ions by the action of an alkali whereas nitrate ions produce the nitrite ions by reduction (e.g. by zinc powder).

The Griess reaction is illustrated in **Fig. 2**.

The reactions mentioned above are the basis of

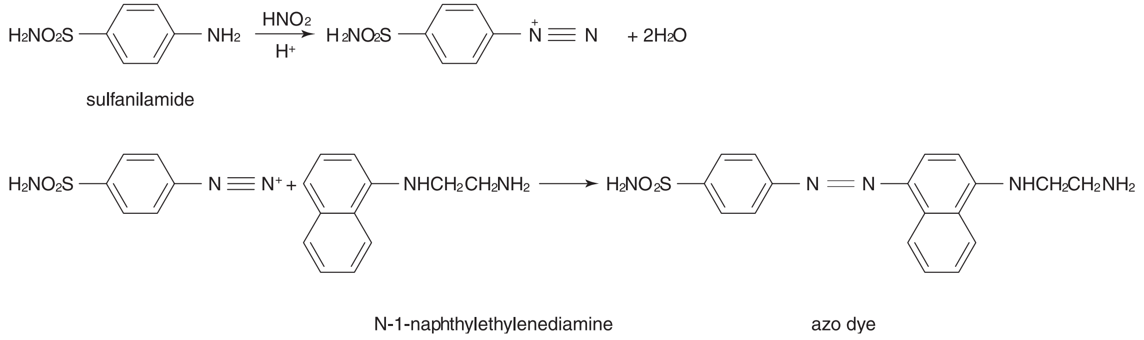


Figure 2 The Griess reaction.

some commercially available kits, often used in border-crossings to screen items or persons suspected of having been in contact with explosives or materials suspected as explosives. An example is the ETK, an Israeli-produced kit, which is capable of detecting some polynitroaromatic compounds, ester nitrates, nitramines and inorganic nitrate salts. The ETK has been successfully used in real cases.

Another type of spot test is based on the oxidation of a reagent by an explosive or an oxidizing constituent of an explosive mixture. Diphenylamine develops a blue color when reacted with TATP; aniline sulfate develops a blue color when reacted with chlorates.

In general, the sensitivity of many spot tests used in the analysis of explosives is in the microgram range. Spot tests are fast, inexpensive, simple, do not need instrumentation and may be performed by technicians in the field. Spot tests cannot be the basis for reliable identification. At best they may strongly indicate that the analyte belongs to a certain class of compounds (e.g. nitroaromatic).

Separation methods Separation methods are used to separate a mixture into its components. Almost all separation methods used nowadays are chromatographic methods. These methods utilize differences in the affinity of the components of a mixture to stationary and mobile phases. Chromatographic behavior of a compound in given stationary and mobile phases is expressed by its migration rate, usually called rate of flow (R_f) or retention time (R_t). The retention data may also depend on other conditions such as temperature (e.g. in GC). Though originally used to separate a mixture, chromatographic methods have often been used to tentatively identify a compound by comparing its retention data to those of a known compound. This is best done by analyzing the analyte and the authentic sample under the same conditions. Chromatographic methods are used in forensic analysis of explosives, both for separation of mixtures and for preliminary identification. Some of the more widely used methods are described below.

TLC In TLC the stationary phase consists of an adsorbent such as silica gel, coated on aluminum or glass plates. Solutions of samples are spotted near one end of the plate which is then placed in a chamber containing the developing solvent (the mobile phase). The developing solvent is allowed to ascend to a certain distance on the plate. Different compounds are carried by the solvent with different rates, resulting in a separation. The plate is then removed from the container and dried. Visualization of the spots is done by ultraviolet (UV) light or by spraying reagents, usually based on the spot tests mentioned above. Both the R_F and the color developed by spraying, if similar to those of an authentic sample, may indicate the identity of a compound. TLC can also be used in a preparative mode in which the separated compound is extracted and its identity confirmed by other methods. The sensitivity of TLC is generally in the microgram to submicrogram range depending on the type of visualization used. TLC is widely used for the analysis of explosives in forensic laboratories.

Although there are many reports describing different TLC systems for the separation of explosive compounds, no single system has been reported to separate all organic explosives. Therefore a combination of several systems is routinely used, such as: (1) 1,2 dichloroethane:acetonitrile (90:10, v/v); (2) trichloroethylene:acetone (80:20, v/v); (3) petrol ether (b.p. 60–80):ethyl acetate (90:10, v/v). **Figure 3** shows some military explosives separated by these TLC systems.

A TLC plate, sprayed with the reagents mentioned above deteriorates quickly; therefore it is recommended that documentation by camera or scanner is carried out immediately after spraying.

TLC is a simple, inexpensive and fast method allowing analysis of several samples in a single run. However, it is considered a low-resolution method and in addition, separation is susceptible to contaminants (when present in high amounts). It must be emphasized that identification cannot be based on TLC only and must be confirmed by other methods.

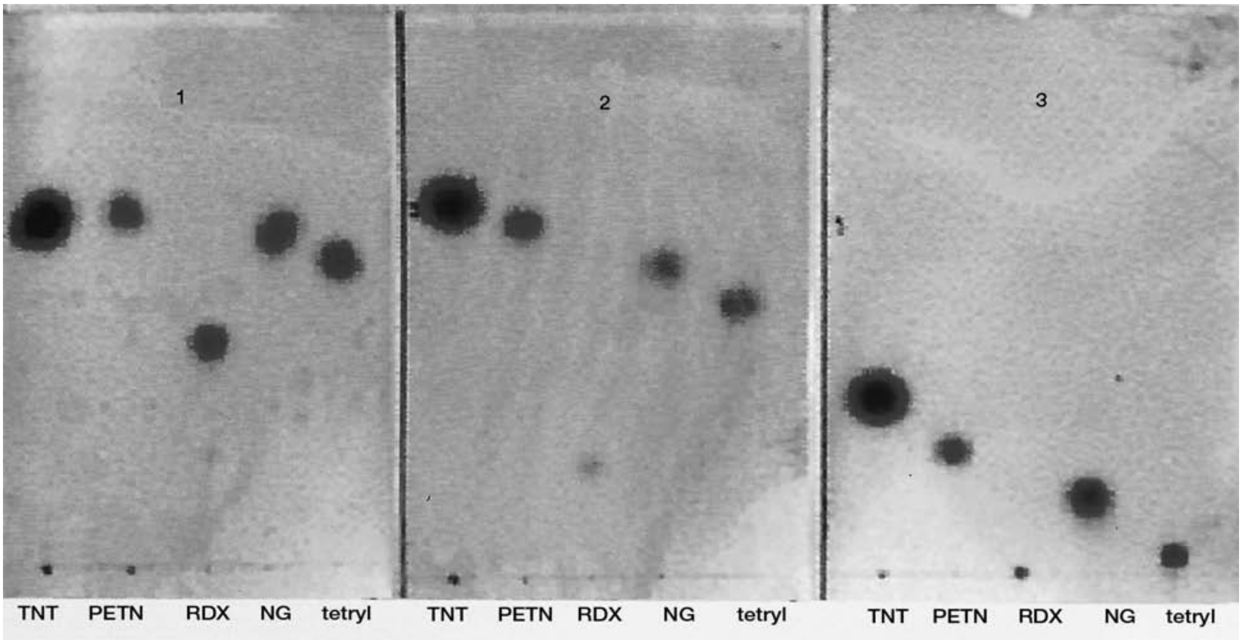


Figure 3 (see color plate 20) Military explosives (TNT, PETN, RDX, NG and tetryl) separated on three TLC solvent systems: (1) 1,2 dichloroethane:acetonitrile (90:10); (2) trichloroethylene:acetone (80:20); (3) petrol ether (b.p. 60–80):ethyl acetate (90:10).

GC Modern GC utilizes capillary columns in which the stationary phase is chemically bonded to the fused silica wall. A sample, dissolved in a suitable solvent, is injected into a heated injection port. The compounds are evaporated and pushed along the coated column, usually at elevated temperatures, towards a detector by the flow of a mobile phase (carrier gas, e.g. helium). Different compounds, with different affinities to the stationary phase, are separated, usually with a good resolution. GC is also a suitable method to perform quantitative analysis.

A typical column used in the analysis of explosives is the nonpolar diphenyl (5%)–dimethylsiloxane (95%) copolymer (e.g. DB-5, HP-5, SP-5), 25 μm coating, 15 m long. Typical GC conditions are: injector temperature 180°C, column temperature programmed from 50°C to 250°C at a rate of 25°C min^{-1}. GC is a simple, high-speed and high-resolution method that is very suitable for the analysis of nitroaromatic explosives. Some loss in sensitivity is observed when analyzing thermally labile compounds (e.g. some nitrate esters and nitramines). Sometimes decomposition products are produced in the injector or in the column. Nonvolatile explosives (e.g. HMX, inorganic compounds) cannot be analyzed by GC. The organic peroxides TATP and HMTD can be analyzed by GC.

Several common detectors are used in GC.

Flame ionization detector (FID) FID, a very common detector in GC used in forensic laboratories, is less common in the analysis of explosives. This is mainly due to a decreased sensitivity towards some explosives which have a high O/C and N/C ratios in their molecules (e.g. NG).

Electron capture detector (ECD) In ECD the eluent from the GC column passes through a slow-electron beam. An analyte containing electronegative atoms such as nitrogen, 'captures' electrons from the constant electron current, producing a signal by decreasing this current. Sensitivity is usually in the picogram range. ECD is quite selective, as it is highly sensitive towards nitrogen-containing compounds but insensitive to hydrocarbons.

Chemiluminescence detector (thermal energy analyser (TEA)) In this detector the eluent from the GC column passes through a furnace which pyrolizes the compounds at elevated temperatures (e.g. 500°–900°C). Nitro and nitroso compounds produce nitrogen oxide which is then allowed to react with ozone to produce nitrogen dioxide in an excited energy level. Decaying of the excited NO_2 to its ground state is accompanied by emission of light at the IR region which is monitored by a suitable detector. Sensitivity is in the picogram range. Commercial Instruments (e.g. 'Thermal Energy Analyser (TEA)') are available and are very sensitive and highly specific for the analysis of nitro-containing compounds (though there have been some reports of signals from compounds without nitro or nitroso groups). GCTEA is suitable for the analysis of postexplosion residues mainly because most contaminants having no nitro

groups, are not observed. GC/TEA is therefore widely used in the forensic analysis of explosives. The TEA detector is expensive, and its use is limited to the analysis of nitro-containing explosives.

MS This is discussed in more detail below.

HPLC In HPLC the stationary phase is often a 'reversed phase' type, such as octadecylsiloxane (ODS). The mobile phase is a solvent or a mixture of solvents which is pumped into the column at relatively high pressure, usually at room temperature. HPLC may be also used in a preparative mode by collecting the desired fraction and confirming its identity by other methods. Some widely used detectors in explosive analysis are discussed below.

Ultraviolet (UV) and diode array (DA) The eluent from the column passes through a cell irradiated by UV light. In the UV detector the source is set at a fixed wavelength or scanned over the UV range. In the DA, multiwave radiation is applied. The UV spectrum obtained by the DA is easily subjected to a computerized library search. These detectors have a low selectivity but their sensitivity is rather high (in the nanogram range) and they are widely used in many forensic laboratories.

Electrochemical detectors In these detectors the detection is based on the electrochemical properties of functional groups in the analyte. A signal is generated when an analyte exchanges electrons with the electrode, if sufficient potential is applied to it. Detection of explosives have been done using pendant mercury drop electrode (PMDE).

Chemiluminescence detector See above.

MS See below.

Ion chromatography (IC) This method utilizes ion-exchange resin as the stationary phase and a solution of salts as the mobile phase. The functional group attached to the stationary phase in anion analysis is usually a quaternary ammonium ion and in cation analysis the exchange function is usually a sulfonate ion. Some instruments utilize a suppressor reaction which takes place in a second column ('suppressor column') situated after the ion-exchange column. This enhances the sensitivity by lowering the background noise of the detector. Sensitivity is in the nanograms range.

IC is widely used in the analysis of explosives and related materials. It is used to separate inorganic ions present in low explosives (e.g. black powder), commercial explosives (e.g. dynamites, ammonium nitrate-based explosives) and home-made explosives (e.g. chlorates).

The common detection modes employed in IC are as follows.

Conductivity detector In this detector a signal is produced when an eluted ion reaches the detector and changes the conductivity of the solution.

UV/VIS Since most inorganic ions do not absorb UV or visible light this detector is used in IC in an indirect mode by adding UV-absorbing eluent to the mobile phase.

Capillary electrophoresis (CE) Chromatography is performed on a capillary column, immersed at its two ends in a buffer solution. SiO^- groups are formed near the capillary wall in the aqueous medium, leading to accumulation of solvated cations. Application of an electric field (e.g. 10–25 kV) on the capillary results in the migration of the solvated cations towards the cathode, generating electro-osmotic flow (EOF). Analytes introduced to the capillary column move in different directions and with different mobilities; negatively charged species move towards the anode, positively charged species move towards the cathode and neutral compounds move with the EOF. Since the EOF is usually faster than the migration velocity of the anions, all species are swept towards the detector which is usually situated near the cathode. Compounds with different mobilities are usually separated with a high efficiency because the profile of the advancing mobile phase is flat, in contrast to its parabolic profile in other separation techniques (e.g. GC and HPLC). Detection is usually by UV where the capillary itself serves as the cell.

In CE it is possible to analyze different classes of compounds: organic and inorganic, neutral and ionic. All these compounds can be analyzed by changing the buffer. Sensitivity is in the picogram range.

The most suitable method for the analysis of ions is capillary zone electrophoresis (CZE) which is applied in some laboratories as a confirmation method for results obtained in IC. **Figure 4** shows the analysis of some ions found in residues of a pipe bomb.

Neutral organic explosives and other related compounds may be separated by micellar elecrokinetic chromatography (MEKC) in which a surfactant (e.g. sodium dodecyl sulfate (SDS)) is added to the buffer, forming micelles. Separation is achieved by partition of the analytes between the micelles and the buffer. The stronger the attraction between the analyte and the micelle the longer is its migration time.

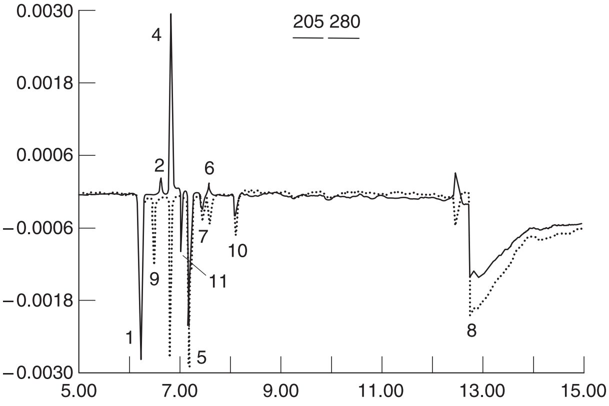


Figure 4 Capillary electropherogram of Pyrodex® pipe bomb residue anions. Analysis was performed at 205 and 280 nm. Peak identification: 1, Cl^-; 2, NO_2^-; 4, NO_3^-; 5, SO_4^{-2}; 6, SCN^-; 7, ClO_4^-; 8, HCO_3^-; 9, HS^-; 10, OCN^-; 11, unknown. Reprinted with permission from McCord BR, Hargadon KA, Hall KE and Burmeister SG (1994) Forensic analysis of explosives using ion chromatographic methods. *Analytica Chimica Acta* **288**: 43–56. With permission from Elsevier Science.

Spectroscopic and spectrometric methods *IR* Molecules irradiated by infrared light (e.g. 4000–500 cm^{-1}) absorb energy at certain wavelengths which correspond to intramolecular vibrations.

Absorbing bands in the range ~4000–1300 cm^{-1} are usually associated with specific functional groups whereas absorption bands below ~1300 cm^{-1} are usually characteristic of the molecule as a whole. Therefore the region below 1300 cm^{-1} is sometimes called the 'fingerprint' region of the IR spectrum. Modern IR instruments, used by most laboratories are Fourier Transform IR (FTIR). IR may be used to identify a pure compound by comparing its spectrum to the spectrum of an authentic sample. Sensitivity is usually in the microgram range; detection limit may be lowered by using microscope FTIR.

Mixtures may require chemical separation of components before the analysis in order to obtain IR spectra of the pure components. Modern instruments usually contain computerized libraries, in which a mathematical 'subtraction' of a spectrum of one component leads to an IR spectrum of a second component. This may reduce the need for chemical separation.

Symmetric and asymmetric stretching vibrations of the NO_2 group give rise to two distinct absorption bands, which have a highly diagnostic value. In nitroaromatic compounds these bands appear at 1390–1320 cm^{-1} and 1590–1510 cm^{-1}, respectively. They can be clearly observed in the IR spectrum of 2,4,6-TNT (**Fig. 5**).

The two NO_2 stretching vibrations in nitrate esters appear at 1285–1270 cm^{-1} and 1660–1640 cm^{-1}, respectively, as can be seen in the IR spectrum of PETN (**Fig. 6**).

The two NO_2 stretching vibrations in nitramines appear at 1310–1270 cm^{-1} and 1590–1530 cm^{-1}, respectively, as can be seen in the IR spectrum of RDX (**Fig. 7**).

An IR spectrum of TATP (which lacks nitro groups) is shown in **Fig. 8**.

Inorganic anions related to explosives also have highly characteristic absorption bands. Nitrate ions absorb at two bands: 1380–1350 cm^{-1} and 840–815 cm^{-1}. Chlorates absorb at 980–910 cm^{-1}, 630–615 cm^{-1} and 510–480 cm^{-1}.

NMR Nuclei, whose nuclear spin is not zero (e.g. ^{1}H, ^{13}C, ^{14}N) behave as small magnets. When

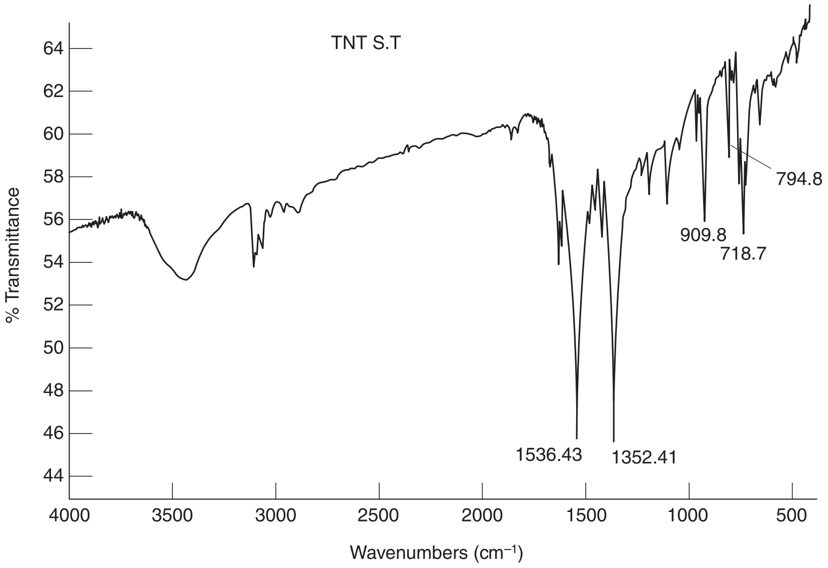


Figure 5 IR spectrum of 2,4,6 TNT.

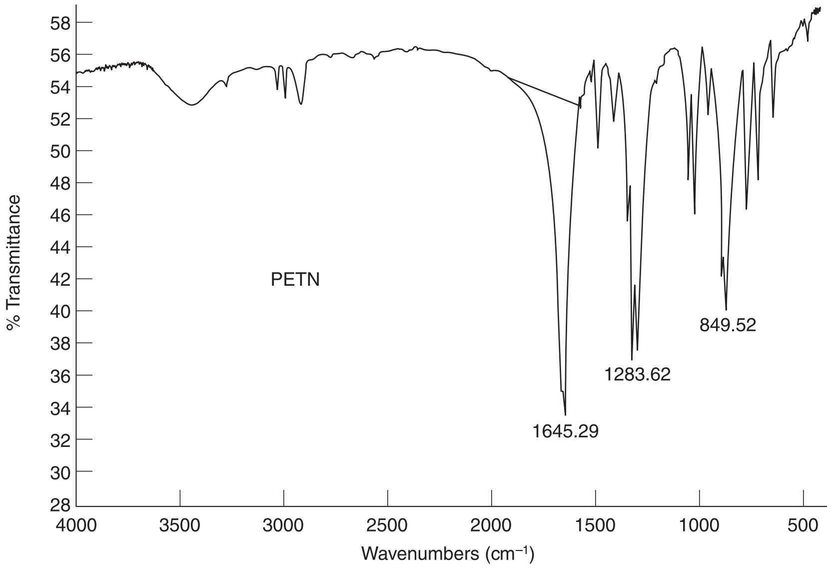


Figure 6 IR spectrum of PETN.

such nuclei are put in an external magnetic field they may align with the magnetic field, having a low-energy orientation, or against it, having a high-energy orientation. Transition between these two energy levels takes place by absorption of suitable radio frequency (RF) radiation called the resonance frequency.

The energy absorbed at such transition depends on the chemical environment of the nucleus; thus, various protons in a molecule resonate at different frequencies. The exact amount of energy absorbed by a specific proton is expressed by its 'chemical shift'. Different protons in the molecule usually have different chemical shifts. Scanning RF while keeping the magnetic field constant, or scanning the magnetic field while keeping the RF constant will result in an NMR spectrum. The NMR spectrum is highly characteristic and may be used for the identification of a compound by comparing its spectrum to that of an authentic sample. NMR is especially useful for structure elucidation of unknown samples even when no authentic sample is available. This is done by correctly interpreting the different signals in the spectrum, leading, often by combination with other methods, to a complete structure elucidation. Most work in NMR has been done on protons producing databases of chemical shifts and spectra. Sensitivity of NMR is

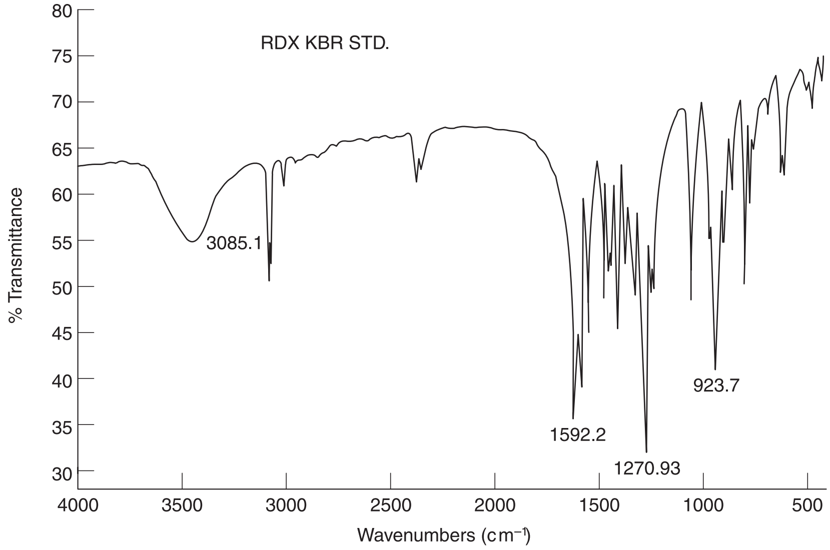


Figure 7 IR spectrum of RDX.

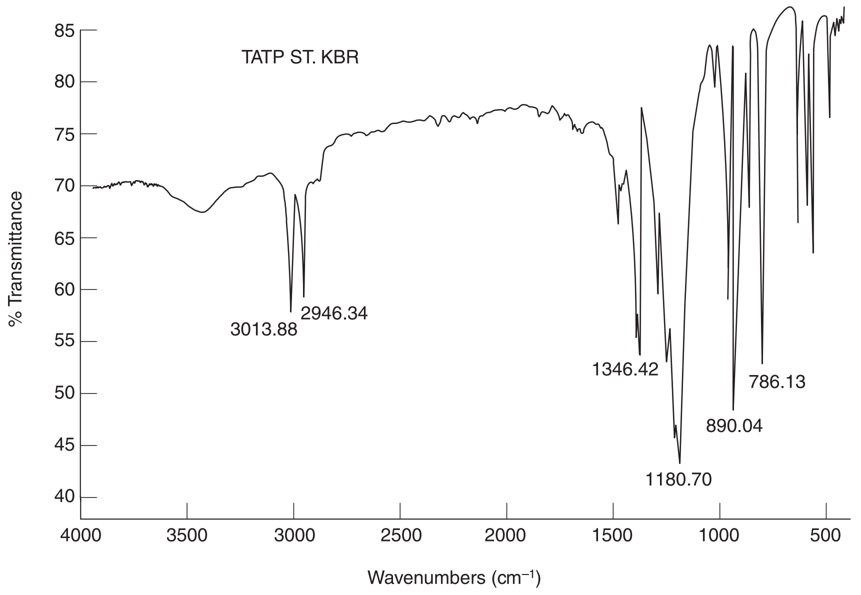


Figure 8 IR spectrum of TATP.

usually in the micrograms to milligrams range. NMR has not been routinely used in the forensic analysis of explosives.

MS In this technique a compound is introduced into an ion source where it is ionized to form molecular and fragment ions according to its structure. The ions pass through an analyzer (e.g. magnet, quadrupole or ion trap) which separates the ions according to their mass to charge ratio (m/z). The ions are detected and recorded, producing a mass spectrum. A mass spectrum often reflects the structure of a molecule. It is usually highly specific and is often referred to as a 'fingerprint' of the molecule. Identification of a compound by its mass spectrum is therefore highly reliable. In addition, a mass spectrum can be used for structure elucidation of unknown compounds. Sensitivity usually lies in the picogram to nanogram range depending on the operation mode.

As it is highly reliable as well as highly sensitive, MS is considered to be an excellent method for the identification of organic compounds.

Introduction techniques include inlets for gases and liquids (usually based on needle valves or gold leaks) and direct insertion probe, also known as 'solid probe' for solids. Solid probe is usually used for the insertion of nonvolatile compounds, such as HMX, and is usually unsuitable for the analysis of mixtures. However, the most common introduction techniques are the on-line combination with GC (GC/MS) and HPLC (LC/MS).

GC/MS The technical possibility of connecting a highly efficient separation method (GC) with a highly sensitive and reliable identification method (MS) was a breakthrough in analytical chemistry. GC/MS enables the separation of highly complex mixtures with the subsequent rapid identification of each of the separated components. Therefore GC/MS is the method of choice in organic analysis in many forensic laboratories. In modern GC/MS instruments, the GC capillary column end is placed near the ion source without the need of an interface, thus enhancing instrument efficiency. As mentioned above some explosives are easily analyzed by GC (hence by GC/MS) whereas with others some difficulties are encountered.

LC/MS requires an interface between high output of liquids and sometimes nonvolatile buffers, and the high vacuum of the MS. This enables analysis of nonvolatile compounds that cannot be analyzed by GC/MS. The common ionization methods in MS are described below.

Electron ionization (EI): In electron ionization, or electron impact, electrons are ejected from a filament and accelerated usually by a voltage of 70 eV. The sample molecules are bombarded by these electrons, to produce mainly positively charged ions. Printed and computerized databases, containing more than 100 000 spectra, enable a useful library search. Attention should be paid by users to limitations of a computerized search such as the following.

- Computerized libraries may contain a partial mass spectrum of a molecule.
- A library search is performed using a certain algorithm. Using different algorithms may result in

different matches for the same compound, which sometimes may lead to erroneous results.

Therefore, the chemist should always exercise judgment and should not rely only on computer results.

Chemical ionization (CI) In this method the ionization of the sample molecules takes place by reactions with ions derived from reagent gases (e.g. methane, isobutane) present in the ion source at relatively high pressure (~1 torr). This method usually produces $[M+H]^+$ ions, where M is a sample molecule and H is the proton transferred from the reagent ions. The $[M+H]^+$ ions have less energy than the molecular ions obtained in EI, resulting in less fragmentation. CI is usually more suitable than EI to obtain molecular weight information and it is best used as a complementary method to EI.

Negative ion mass spectrometry Although less common than mass spectrometry of positive ions, this has some relevance to the analysis of explosives. Formation of negative molecular ions in EI is usually less efficient than formation of the positive ions. In addition, the sensitivity in negative ion MS often depends on the electron affinity of the atoms. CI may also be used in the negative mode, often producing ions which are characteristic of the sample molecules.

Thermospray ionization This term describes both an inlet system and an ionization method for HPLC/MS. The eluent from the HPLC is rapidly vaporized in a heated metal capillary tube, forming fine droplets. These droplets are ionized by a filament or by electrolyte ions present in the solution. An analyte ion may be generated in the solution and transferred to the gas phase, or generated in the gas phase by ion–molecule reactions. In this method positive and negative ions are formed.

Electrospray ionization (ESI) This is an ionization technique for HPLC/MS or CE/MS which takes place at atmospheric pressure. The eluent from the separation device passes through a stainless-steel capillary to which high positive (or negative) potential (e.g. 3–5 kV) is applied. This potential leads to the formation of charged droplets. Evaporation of the solvent from the droplets (e.g. by a stream of gas) leaves behind a charged analyte.

Mass spectra Mass spectra of common explosives are discussed below. EI mass spectra of nitroaromatic compounds are highly specific, enabling the differentiation of isomers. A loss of NO_2 groups from the molecular ion is an important process, leading to characteristic $[M\text{-}xNO_2]^+$ ions. Unlike nitrate esters, the abundance of NO_2^+ ions in the spectra of nitroaromatic compounds is low. The abundant ion ('base peak') in the EI spectrum of 2,4,6-TNT (m/z 210) is attributed to an unusual loss of a hydroxyl group from the molecular ion. This loss also occurs in the EI spectrum of other nitroaromatic compounds with a nitro group in an ortho position to a hydrogen-containing moiety such as methyl ('ortho effect'). The molecular ion is not observed in the EI mass spectrum of 2,4,6-TNT as shown in **Fig. 9**.

EI spectra of nitrate esters contain highly characteristic ions at m/z 30, 46 and 76, attributed to $[NO]^+$, $[NO_2]^+$ and $[CH_2ONO_2]^+$, respectively. No molecular ions are observed in the EI spectra of these compounds, so the EI spectra of most common nitrate ester explosives are similar, showing ions only at m/z 30, 46 and 76. Therefore, the EI spectrum alone is not sufficient for identification of nitrate esters. CI may

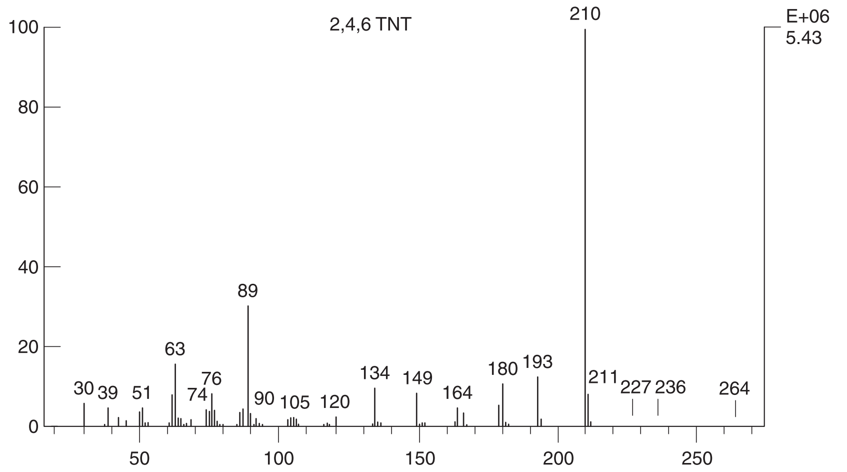


Figure 9 EI mass spectrum of 2,4,6 TNT

serve as a very good confirmation method for these compounds. Their CI spectra contain not only molecular weight information (e.g. $[M+H]^+$ ion), but also characteristic structural information: an abundant ion at m/z $[M+H-63]^+$ is attributed to the loss of $HONO_2$ from the protonated molecule. EI and CI mass spectra of the two nitrate esters, NG and EGDN, are shown in **Fig. 10**.

EI spectra of nitramines usually lack the molecular ion. The two heterocyclic nitramines RDX and HMX have characteristic EI spectra on which unequivocal identification can be based. The EI spectrum of RDX is shown in **Fig. 11**.

Tetryl, a nitramine with a trinitroaromatic nucleus, decomposes in the GC injector in the presence of water, to produce *N*-methylpicramide. The reaction is shown in **Fig. 12**. Therefore, the original presence of tetryl should be confirmed by other methods such as IR or TLC which are carried out at room temperature. The EI mass spectrum of *N*-methylpicramide is shown in **Fig. 13**.

The EI spectrum of TATP (**Fig. 14**) contains a low abundant molecular ion at m/z 222.

Other compounds such as plasticizers and stabilizers are usually easily analyzed by GC/MS.

XRD X-ray diffraction, or X-ray powder diffraction (XRPD), utilizes X-ray radiation on crystalline organic and inorganic samples. The rays are diffracted in a pattern determined by the position, arrangement and size of the constituents of the crystal. Scattered photons, which may undergo subsequent interference, lead to a characteristic diffraction pattern. The pattern is characteristic for a specific crystalline powder and may serve as a 'fingerprint' of this powder. Identification of a powder may be carried out by comparing its spectrum to a spectrum of an authentic sample. Data bases of diffractograms are commercially available and it is also possible to analyze and identify multiphase crystalline mixtures, qualitatively and quantitatively. Sensitivity is usually in the microgram to milligram range.

XRD is mainly used to analyze crystalline powders of inorganic explosives and explosive mixtures. The advantage of this method over IC or CE is the identification of a compound as a whole; in the

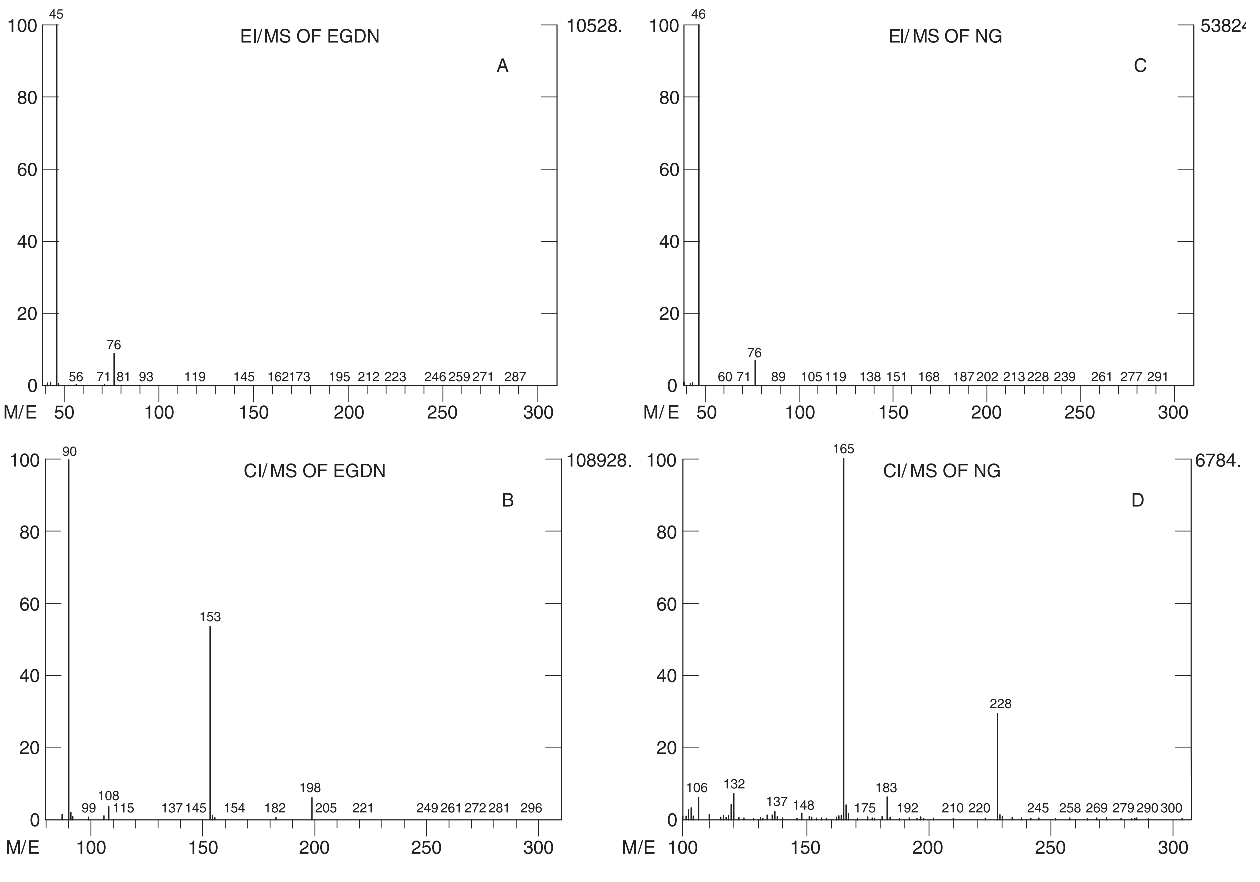


Figure 10 EI mass spectrum (A) and CI mass spectrum **(B)** of EGDN. EI mass spectrum **(C)** and CI mass spectrum **(D)** of NG. Reprinted from Tamiri T and Zitrin S (1986) Capillary column gas chromatography/mass spectromentry of explosives. *Journal of Energetic Materials* **4:** 215, with permission from Dowden Brodman and Devine, Inc.

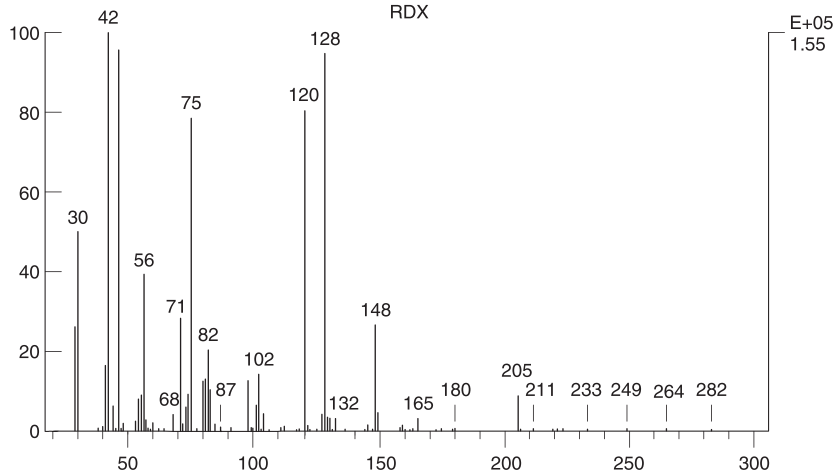


Figure 11 EI mass spectrum of RDX.

latter methods anions and cations are identified separately.

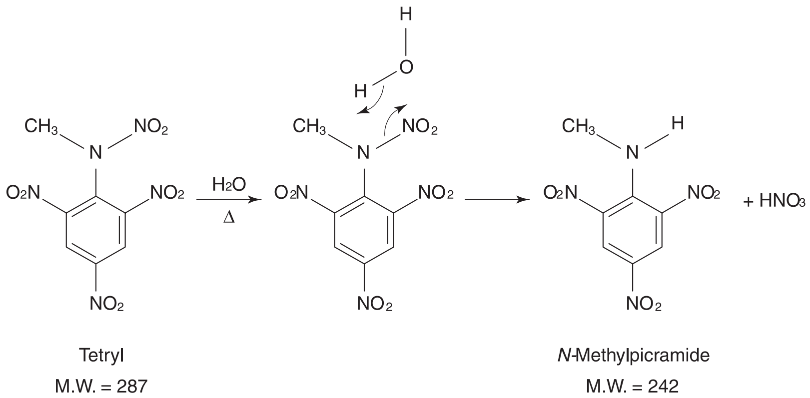
XRD instruments are expensive and require expert operators.

SEM/EDX This technique enables examination and morphological characterization of surfaces of organic and inorganic samples. The sample is bombarded by a high-voltage (e.g. 25 kV) electron beam. An interaction between the sample and the electron beam causes emission of radiation in the X-ray range typical of an element. EDX permits high-speed qualitative and quantitative elemental analysis according to the intensity of the energy emitted by the sample. Elements with the atomic number 11 (sodium) and higher may be analyzed by this technique. Special light-element detectors enable the identification of elements with the atomic number 5 (boron) and higher. This has relevance to the identification of explosive compounds which often contain nitrogen atoms. SEM/EDX is suitable for the identification of metals present in primary explosives such as lead azide or mercury fulminate.

SEM/EDX instruments are expensive and require expert operators.

Postexplosion and Trace Analysis of Explosives

Postexplosion analysis is normally based on the identification of the unreacted explosive which 'survived' the explosion. Usually only trace amounts of the unexploded explosives are present. The other alternative, i.e. analyzing decomposition products which are typical of the original explosive, is only carried out in rare cases. An example is the thiocyanate ion (CNS^-) whose formation is characteristic of the burning of black powder.



> **Figure 12** Hydrolysis of tetryl to *N*-methylpicramide. Reprinted from Tamiri T and Zitrin S (1986) Capillary column gas chromatography/mass spectrometry of explosives. *Journal of Energetic Materials* **4:**

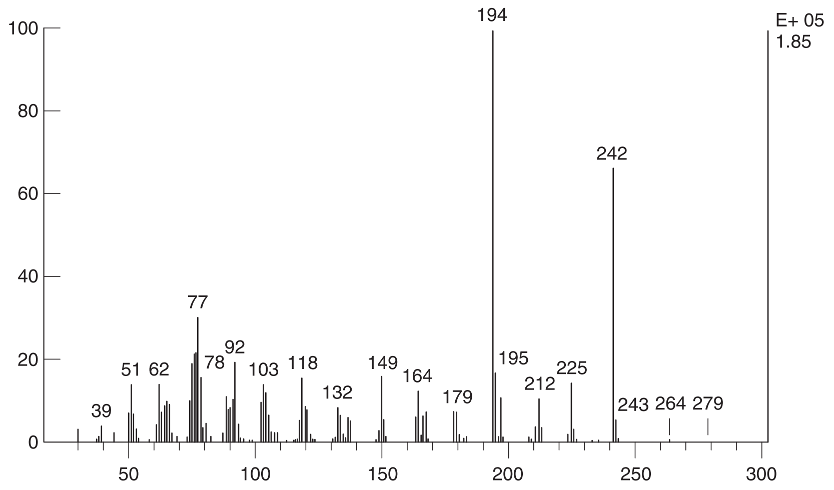


Figure 13 EI mass spectrum of N-methylpicramide.

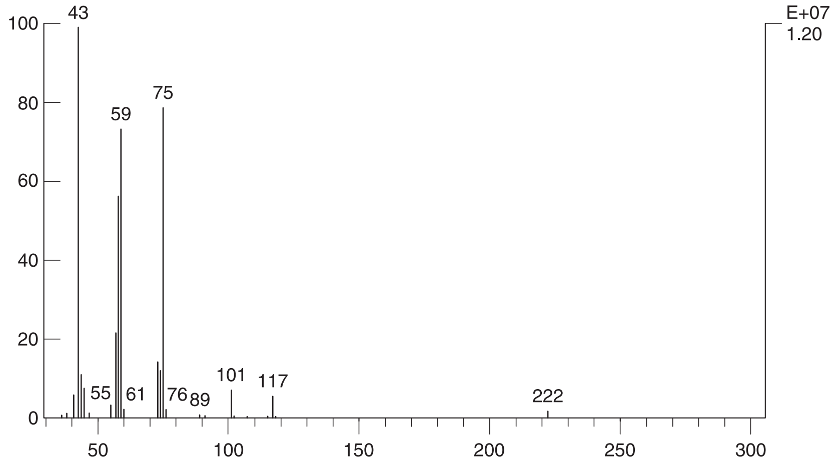


Figure 14 EI mass spectrum of TATP.

Analysis procedures

A typical flow chart for postexplosion analysis includes sampling, visual examination, vapor analysis, extraction procedures and analysis by various methods. Extraction procedures include extraction by organic solvents for the organic explosives and aqueous extraction for the water-soluble compounds. Cleaning procedures may be carried out prior to the analysis, if required.

A flow chart of the generalized analysis of postexplosion is given in **Fig. 15**. The expert may choose a suitable debris procedure and an appropriate identification method according to the circumstances.

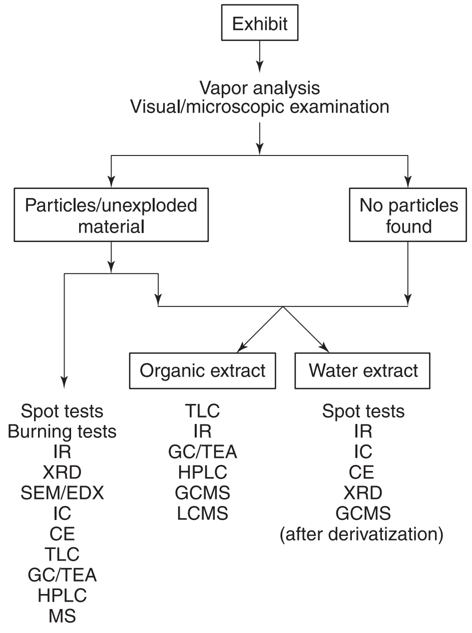


Figure 15 Flowchart of analysis of postexplosion debris

Sampling methods

Modern analytical methods can detect extremely small amounts of material, but no method, however sensitive, can detect an explosive if it is not present on the exhibit. Therefore, the collection of exhibits is a

critical step in postexplosion analysis. Much work has been done to establish a methodology for the collection of exhibits according to their distribution around the bomb site. Unfortunately, no such methodology has been proved to be efficient and it seems that luck plays an important role in collecting the 'right' exhibit. Attempts to overcome this problem have been made by screening exhibits at the explosion scene, using kits based on color tests or 'sniffing' devices. Sniffing instruments are usually based on chemiluminescence detection (e.g. EGIS®) or ion mobility spectrometry (IMS) (e.g. IonScan®). The results of these preliminary tests are of indicative value only and cannot be regarded as an identification.

A major problem in postexplosion work is the potential contamination, either at the stage of collecting exhibits or during the subsequent laboratory analysis. Anyone collecting exhibits should also submit swabs of himself, taken before the sampling, to reveal the possibility of cross contamination. These swabs should be sent to the laboratory, considered as 'blank samples', and processed by the same procedures used for the exhibits.

Visual examination

It is highly recommended that the analysis of an exhibit begins with visual examination. As is usual in routine forensic work, nondestructive methods should be used first in order to extract information that may be lost after the use of destructive methods. Such information may include the morphological appearance of a particle which may connect a suspect to the scene if an identical particle is found in a suspect's possession. The naked eye or a low-power stereoscope may be used for the physical separation of particles such as black powder, smokeless powder or material not consumed in the blast. Sieving may also be used for separation of these particles from debris.

Vapor analysis and adsorption on solid phase

Volatile explosives such as TATP or NG may be detected in vapors of an exhibit by direct headspace analysis. Adsorption of explosives from the vapor phase may be carried out at the scene (or in the laboratory) by passing the vapors through a suitable adsorbent material such as Amberlite XAD-7® or Tenax®. Explosives adsorbed on such resins may be eluted by a suitable solvent and then analyzed.

Fig. 16 shows mass chromatograms and total ion current (TIC) in which TATP was identified in vapors of soil, taken from a real-life case.

Organic extraction

Organic extraction is usually performed with acetone, which is the most commonly used solvent for explosive compounds. The solvent is then evaporated, under a stream of nitrogen rather than by heating, in order to minimize evaporation of volatile explosives. Acetone also dissolves nonexplosive materials from the debris such as oily compounds (e.g. hydrocarbons, fatty acids), plasticizers (e.g. phthalates) and some polar compounds. These materials may coelute with the explosives in chromatography and even change the R_t of the explosives. These coeluting compounds may also cause a significant decrease in instrumental performance. For example, contamination of the injection port, column and ion source in GC/MS results in decrease in sensitivity and resolution. To minimize extraction of oily compounds, ethanol/water mixtures rather than acetone may be used to swab exhibits.

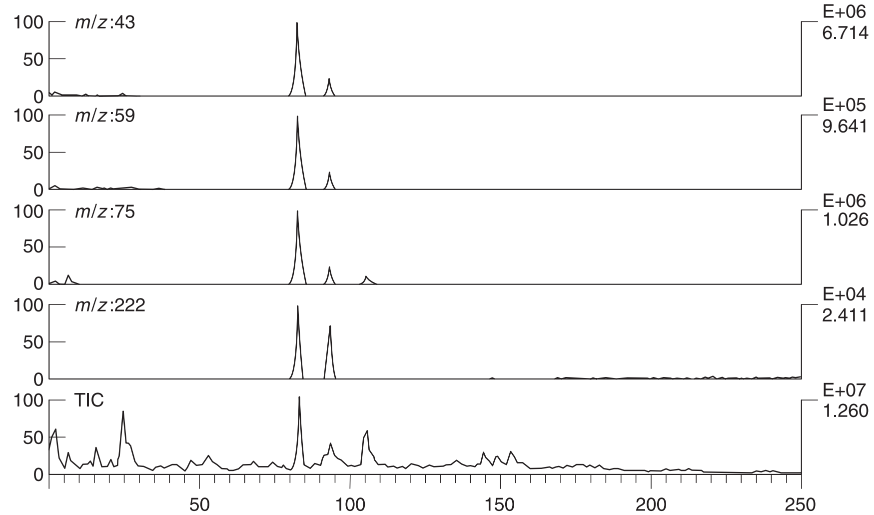


Figure 16 Total ion current (TIC) and mass chromatograms at *m/z* 43, 59, 75 and 222, of soil taken from a real-life case. Vapors of soil were adsorbed on active charcoal and eluted by CS_2. TATP was identified in the peak emerging after 83 s.

Cleaning procedures In order to reduce the amounts of contaminants in the extract, cleaning procedures may be carried out prior to the analysis. They include liquid/liquid extraction, preparative TLC or HPLC, and solid phase extraction (SPE). SPE employs a suitable adsorbent packed in a column or in commercially-available cartridges. The extract is mounted on the adsorbent; starting with nonpolar eluting solvents (e.g. hexane), the hydrophobic compounds (e.g. hydrocarbons) are washed out first and the explosives are eluted later, when more polar solvents are used.

Analysis Analysis of a completely unknown compound often starts with screening tests (e.g. TLC, HPLC) followed by a suitable confirmation method (e.g. MS). In addition to serving as a screening method, GC/TEA, being highly specific, may also serve as a confirmation method. An advantage of using TEA detection rather than UV detection in postexplosion analysis is demonstrated in **Fig. 17**. A 'real sample' with traces of PETN was analyzed by HPLC with UV at 210 nm (upper trace) and by GC/TEA (lower trace). PETN could not be unequivocally identified in HPLC/UV chromatogram whereas the chromatogram obtained by GC/TEA could be safely interpreted.

The forensic explosive laboratory at the Defence Evaluation and Research Agency (DERA) in the UK performs three consecutive GC/TEA analyses using three different columns. An identification of an explosive is concluded only when all three analyses are positive for the compound. GC/MS has been successfully used in the analysis of organic extracts of post explosion debris. **Fig. 18** shows the TIC and mass chromatogram of the ion at m/z 46 where RDX was identified in the extract of postexplosion debris from a real-life case.

Some work has been carried out using negative ion mass spectrometry. IR is usually not suitable for postexplosion analysis due to the large amount of

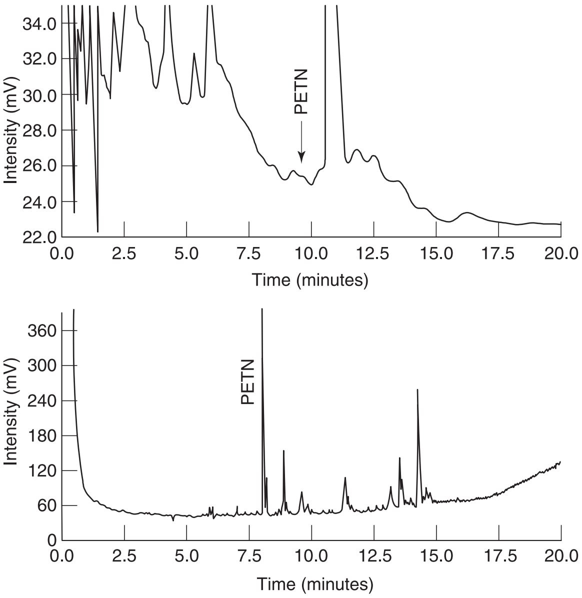


Figure 17 Upper trace: HPLC chromatogram with UV detection at 210 nm of a 'real sample' with traces of PETN. Lower trace: chromatogram of the same sample analyzed by GC/TEA. Reprinted from Kolla P (1991) Trace analysis of explosives from complex mixtures with sample pretreatment and selective detection. *Journal of Forensic Science* **36:** 1342, with permission from ASTM.

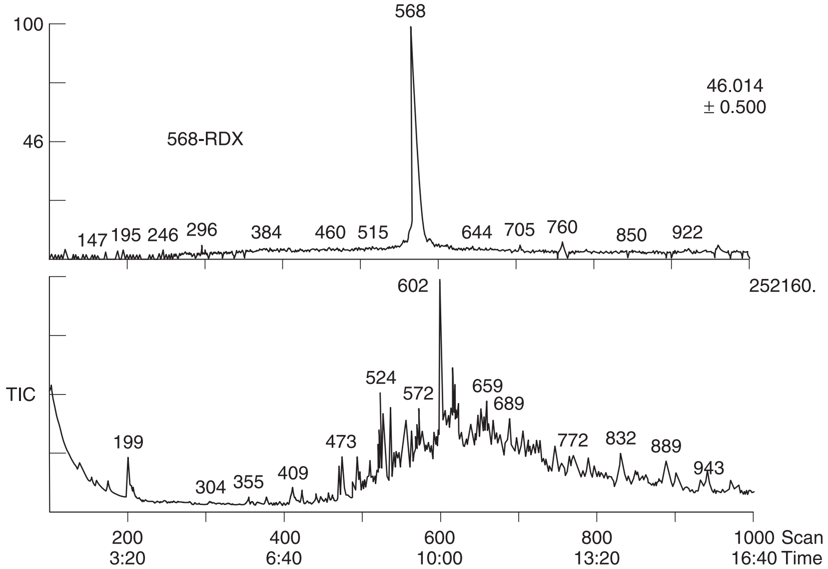


Figure 18 TIC and mass chromatogram at ion *m/z* 46, of postexplosion debris from real-life case. RDX was identified in the peak emerging after 568 s.

contaminants which interfere with the spectrum. The application of NMR to postexplosion extracts has had limited success, mainly due to its low sensitivity.

Aqueous extraction and analysis

Water is used to extract hydrophilic, water-soluble compounds. The dried extracts are then subjected to further analysis. Inorganic explosive-related anions may be detected by spot tests and confirmed by IC or CE. IR may also be used for the identification of these anions and other water-soluble compounds such as sugars. Unequivocal identification of some explosive-related inorganic anions may be carried out by GC/MS. Nitrate, nitrite, thiocyanate and sulfide anions are derivatized by a suitable reagent such as pentafluorobenzylbromide (PFBB), to produce volatile compounds which are easily analyzed by GC/MS. **Figs 19** and **20** show EI mass spectra of the derivatization product of nitrate and thiocyanate, respectively. This method has been successfully applied to real-life casework.

4. Criteria for Identification

In forensic analysis where the results may affect the verdict in a court of law, a mistake must be avoided at all costs. Hence an identification of an explosive should be concluded only if strict criteria are met. It is generally accepted that unequivocal identification cannot be based on spot tests alone, as the results are presumptive and can at best indicate the possible presence of the chemical class to which the explosive belongs. The situation is more complex with chromatographic methods. The conservative view has been that the identification of a single organic compound could not be based on chromatographic methods alone even if a combination of several such methods was used. Results from methods such as IR, MS or NMR are also required, because they reflect properties which are more directly related to the molecular structure of the analyte than chromatographic methods. This criterion has posed difficult problems in postexplosion analysis. The chromatographic results, which indicated the presence of an explosive, could not always be confirmed by such

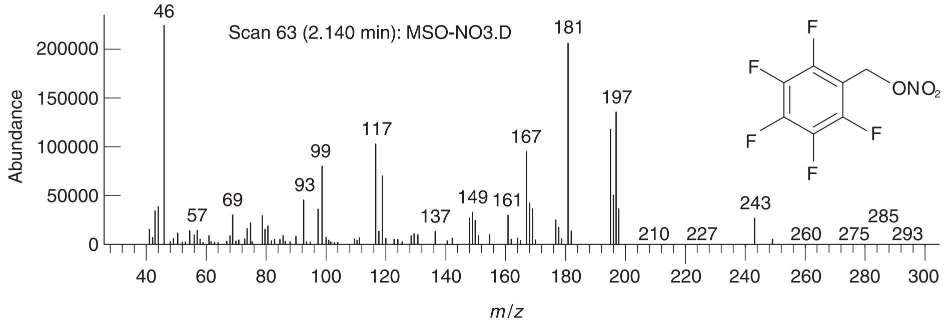


Figure 19 EI mass spectrum of the derivatization product of nitrate.

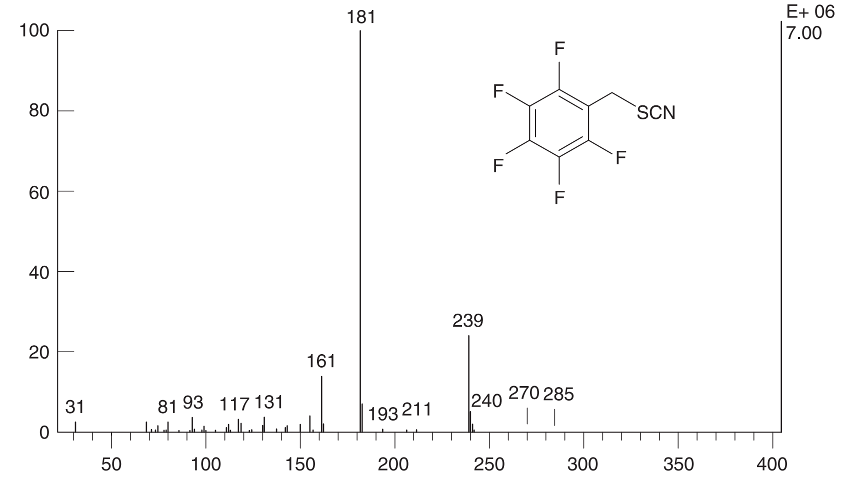


Figure 20 EI mass spectrum of the derivatization product of thiocyanate with PFBB.

methods because of the extremely small amounts of unreacted explosives mixed with the large amounts of contaminants.

This view is currently being re-examined by some forensic analysts. It can be argued that the introduction of highly specific detectors to chromatographic methods (mainly GC and LC) could change the criteria in such a way that a positive result in a chromatographic analysis could suffice for unequivocal identification of an analyte. The most striking example is the introduction of the TEA for GC and HPLC analysis of nitro-containing explosives. An indispensable prerequisite for a positive response of the TEA is the presence of nitro (or nitroso) group. Without their presence there would be no response. This enhanced specificity has led forensic analysts to regard chromatographic retention data combined with a positive TEA response as sufficient evidence for the identification of nitro-containing explosives.

As can be seen in this chapter to obtain optimum results in the analysis of explosives in forensic laboratories the most modern and sophisticated instrumentation should be used. The human factor, however, remains the most important element in forensic science. It is the skill, experience and intelligent judgment of the chemist which determine the quality of the analytical results.

See also: **Analytical Techniques**: Separation Techniques; Spectroscopy: Basic Principles; Presumptive Chemical Tests; Mass Spectrometry; Hyphenated Chromatographic-Spectroscopic Techniques. **Explosives**: Bomb-scene Management.

Further Reading

Beveridge AD (1992) Development in the detection and identification of explosives residues. *Forensic Sciences Review* **4**:18–49.

Beveridge AD (ed.) (1998) *Forensic Investigation of Explosions*. London: Taylor and Francis.

Chasan DE and Norwitz G (1972) Qualitative analysis of primers, tracers, boosters, and delay compositions on a microscale by use of infrared spectroscopy. *Microchemical Journal* **17**:1–60.

Fedoroff BT, Aaronson HA, Reese EF, Sheffield OE and Clift GD (1960) *Encyclopedia of Explosives and Related Items*, vol. 1, PATR 2700. Dover, NJ: Picatinny Arsenal.

Fetterolf DD (1995) Detection and identification of explosives by mass spectrometry. In: Yinon J (ed.) *Forensic Applications of Mass Spectrometry*, pp. 215–257. Boca Raton: CRC Press.

ICDD (JCPDS) (1995) PDF-2 Data base (1995), 12 Campus Boulevard, Newton Square, PA (1995), USA.

Lloyd JBF (1992) HPLC of explosives materials. *Advances in Chromatography* **32**:173–261.

McCord BR, Hargadon KA, Hall KE and Burmeister SG (1994) Forensic analysis of explosives using ion chromatographic methods. *Analytica Chimica Acta* **288**:43–56.

NIST/EPA/NIH mass spectral library Version 1.0 (1995) US Department of Commerce, Technology Administration, NIST, Standard Reference Data Program, Gaithersburg, MD.

Nyquist RA and Kagel RO (1971) *Infrared Spectra of Inorganic Compounds (3800–45 cm^{-1})*. New York: Academic Press.

Pristera F, Halik M, Castelli A and Fredericks W (1960) Analysis of explosives using infrared spectroscopy. *Analytical Chemistry* **32**:495–508.

Saferstein R (1998) *Criminalistics, An Introduction to Forensic Science*, 6th edn, pp. 343–360. Englewood Cliffs, NJ: Prentice-Hall.

Urbanski T (1964) *Chemistry and Technology of Explosives*, Vols 1–3. Oxford: Pergamon Press.

Yinon J and Zitrin S (1981) *The Analysis of Explosives*. Oxford: Pergamon Press.

Yinon J and Zitrin S (1993) *Modern Methods and Applications in Analysis of Explosives*. Chichester: John Wiley.

Bomb-scene Management

P Meikle, New Scotland Yard, London, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0511

'The basic objectives of a bomb-scene manager are to facillitate the recovery of evidence and return the scene to the public domain as soon as practical. Within the London area the bomb-scene manager is a Detective Inspector from the Anti Terrorist Branch of New Scotland Yard who has received specific training in this field.

In the event of a bomb exploding within the Metropolitan Police area the first response will always be by uniformed members of the three emergency services, whose primary duties are preserving life, treating and removing the injured and extinguishing fires. Little can be done at this stage to preserve the scene until the withdrawal of the emergency services upon completion of their core functions; however, there is a growing program of communication whereby each organization is becoming more aware of the need to

limit the disturbance of any bomb scene prior to the forensic examination.

Having been notified of a bomb explosion, the bomb-scene manager will attend the site, accompanied by an exhibits officer, and determine the best deployment of cordons. The cordons are essential to prevent any further disturbance to the scene and to ensure public safety. After discussion between the bomb-scene manager and the senior uniformed officer present, two cordons are deployed – outer and inner; their position is dictated by the geography of the scene. The inner cordon is placed beyond the limits of the explosion as determined by the blast and debris. Only the bomb-scene manager and his team of officers may enter this area until the completion of their forensic examination and evidence retrieval. The outer cordon is placed some distance from the inner cordon, allowing a safe working area for police and other members of the emergency services. The outer cordon is manned continuously by uniformed police officers to ensure security of the scene. At the same time the bomb-scene manager will speak directly to the explosives officer present to establish that the area is 'safe' and establish any views or opinions that the officer may have, particularly with regard to cordons and safety. Outside London, this role is performed by members of the 11th Ordnance Battalion (EOD), who provide 24-hour cover. London is particularly fortunate in that the Metropolitan Police employs 12 ex-members of 11 EOD to fulfill this role. They also provide 24-hour cover and invaluable advice and assistance to the bomb-scene manager.

The exhibits officer will then evaluate the scene and determine appropriate paths of entry beyond the inner cordon; a police photographer will be summoned before any physical examination takes place. The photographs taken are extensive; the views are dictated by the exhibits officer. This process takes time, which allows the officers to plan what resources, in terms both of materiel and personnel, will be required to progress the forensic retrieval.

It is standard practice for all officers deployed as members of the forensic recovery team to wear protective clothing, hard hat, gloves and reinforced boots. The protective clothing is donned at the point of entry through the inner cordon and removed at the same point when leaving. This serves two main purposes. First, it provides a uniform standard of dress for all officers, making them identifiable to their uniformed colleagues manning the cordons and providing basic safety standards. Second, it ensures that no contaminants are accidently taken into the scene or 'walked' out beyond the inner cordon.

The investigation into the offense will have begun, a senior investigating officer appointed and the bomb-scene manager will liase with him or her, advising details of the situation and any avenues of investigation that can be undertaken immediately. Having reviewed the scene, the scene manager will then call for the appropriate personnel, a control van and equipment vehicle. Both these vehicles pass through the outer cordon to a convenient point at or near the inner cordon, along with vehicles that the team arrive in. On no account do they enter the area of the inner cordon. The team of officers will be fully briefed by the bomb-scene manager and, depending on the scene, one officer may be deployed in the control van to ensure permanent communications, to act as an immediate point of contact for other officers and to maintain a log of all events and persons entering and leaving. At this stage all deployed officers will don their protective clothing, which is kept for immediate use at New Scotland Yard.

Consultation also takes place with the Fire and Ambulance Services as to any hazards that may exist and any precautions which should be taken. It is common practice at the scene of a major incident for both of these services to leave one unit on scene after the bulk of personnel withdraw, as a safety back-up for those investigating the incident. This is an invaluable resource at the beginning of any investigation where there is a possible hazard to the officers deployed. When the scene is considered safe by the bomb-scene manager, and after consultation with these units, they also withdraw to their other duties. During any standby period the vehicles of these two services remain between the outer and inner cordons.

The safety and welfare of the forensic recovery team is the responsibility of the bomb-scene manager, who must carefully consider the advice given by the other rescue services. The main hazards arise from dangerous structures caused by the bomb blast and falling debris, particularly glass. Also to be considered are noxious gases and dust. In the event of the scene being severly polluted by air-borne debris, masks with filters can be supplied at short notice, allowing the forensic retrieval team to advance. In some cases additional advice with regards to health and safety may be sought from other agencies. In any large city, falling glass is an ever-present danger and, as a standard precaution, one or more officers are deployed as lookouts with the sole purpose of alerting the forensic team working in a potentially hazardous environment.

The bomb-scene manager and the exhibits officer will take an overview of the scene to determine if any immediate avenues of investigation are available. In particular, the scene will be examined to ascertain if the area and its surrounds are monitored by closed-circuit television (CCTV). If so, steps would be taken

to retrieve the video, if one exists, in order that it can be evidentially examined by officers at New Scotland Yard.

Consultation and liaison play a major part during the bomb-scene manager's deployment at the scene of a bomb explosion. The local police commander, either directly or through a contact officer, will be advised of the extent of the damage and the length of time the examination of the scene may take. This is vital in order that he or she can advise the local community accordingly, to minimize public concern. Consideration also needs to be given to disruption to the local community, diversion of bus routes, resulting traffic congestion and possible station closures, if any are within the cordoned area. If it is necessary to close a railway station, the bomb-scene manager would liase direct with the British Transport Police in order that they can issue notices and advise the appropriate railway operators.

Members of the media will now have arrived at the scene in large numbers and will be held in a designated area outside the outer cordon. On the whole they will respect cordons and police instructions in a good-natured manner, but, as they work to restrictive deadlines, pressure is placed on the local police commander for information and photographs. Where possible, the bomb-scene manager, together with the exhibits officer, will provide a route of access whereby nominated members of the media can be escorted to a vantage point in order to film or take photographs. Having done so and been briefed by the police commander, the media representatives will invariably withdraw until the following day, when this exercise may have to be repeated. This is also a useful opportunity for the police to consider making public appeals through the media.

The next stage in the management of the scene is to zone the inner cordon. The zones depend on the geography of the area and the spread of debris caused by the blast, and are defined on a map drawn by the exhibits officer. The zoned areas are then 'white taped' within the inner cordon and designated by number. The purpose of zoning is to ensure that every part of the scene is meticulously examined using a methodical approach, and only on completion of the examination and clearance of the first nominated zoned area will the forensic team move to the next.

If, during the initial rescue phase, bodies are discovered, they are left *in situ* for the forensic team to recover. Where the body or body part lies in a sheltered area, it should be left uncovered; however, if within the public view, even if at a distance, it is normal practice to cover the body with whatever material is to hand. Life must be pronounced extinct by a medical practitioner, even in the most obvious cases, and details of the doctor and the time life was pronounced extinct are relayed to the bomb-scene manager. Taking into account the distress such a fatality can cause to family and friends, steps are taken to recover the body as soon as it is practical to do so, and in such a manner as to accord the deceased such dignity as the circumstances allow. At this stage there should be close liaison with the local coroner, who may wish to examine the scene and the body *in situ*. This is normally arranged through the coroner's officer.

Having established the location of the body, bodies or body parts within the zoned areas, the exhibits officer will clear an approach route to the body, white taping the cleared route for the assisting officers to follow when called for. Before the body is touched, a photographer is summoned and the body photographed in detail. The photographer is provided with protective clothing and footwear which he or she dons before entering the inner cordon. Having taken the photographs, which may include polaroids, the photographer will leave the scene using the cleared approach/exit path and remove the protective clothing when leaving the inner cordon. If life has not been pronounced extinct, or it cannot be ascertained when and by whom, a doctor will be called and escorted to the body by the approach path. Once again the doctor will be provided with footwear and overalls before entering the inner cordon, and removes them on leaving. During this process the exhibits officer makes careful notes of what takes place. If the body has been covered, the body will first be photographed as found. The covering will then be removed by the exhibits officer, sealed in a bag and identified with a unique number. The body is then photographed further, particular attention being paid to the face, injuries and overall position of the body. If the body is covered in debris, this is gradually removed, the body being photographed at each stage.

The exhibits officer will then examine the body, ensuring that no items of visible evidence are apparent, and then proceed to 'bag' the body before it is removed. While in most cases the identity of the deceased is known, or believed known, it cannot be assumed that the deceased is not the person responsible for, or connected with, the explosion and therefore steps are taken to ensure that any evidence is preserved and contained with the body. In order to do so the exhibits officer will separately bag both hands, both feet and the head of the deceased in nylon bags sealed with tape. The entire body is then wrapped in a plastic sheet, which is sealed. The body is then placed in a body bag and removed from the scene. This method of 'bagging and sealing' ensures

that no evidence or contaminants are lost from the body during removal. Each stage of this process is carefully noted by the exhibits officer and the body identified by number. When the body is removed, the debris from under the deceased will be collected by the exhibits officer, carefully placed in a bag and logged. Further bodies are dealt with in an identical manner. In the case of body parts, such as arms and legs, there is no requirement for a doctor to pronounce life extinct, otherwise the procedure is the same. In situations where a body has been severed but the parts remain in close proximity, they must be treated separately. Identification of which body parts belong to which body is a matter for the pathologist.

The body, once removed from the scene, is accompanied by an appointed officer to a designated mortuary. Before any postmortem examination takes place, the body is X-rayed to determine the location of any fragments or splinters within the body which may be linked directly to the bomb; these are of vital importance to the investigating officers. The same officer will identify the body to the pathologist and be present throughout the postmortem examination to receive any items of evidential value. A procedure exists for insuring that the integrity of the exhibits is maintained.

During the investigation, one or two officers with specific skills and knowledge compile a detailed map of the bomb-scene area, on which will be plotted evidential finds. The map shows in detail debris, the positioning of vehicles, bodies and any other physical or geographic areas that may have some relevance. These officers also have the responsibility of preparing a briefing note on the incident for other interested agencies and identifying as rapidly as possible the nature of the device and its component parts. This is particularly important in efforts to 'backtrack' the investigation to the purchase or acquisition of each part of the bomb, which may lead to the apprehension of those responsible.

Depending on the location and the sensitivity of the investigation, screens can be called for; these can be rapidly erected and allow the forensic recovery team to perform their duties with some degree of privacy. The bomb-scene manager will also consider the use of tents to protect the scene from inclement weather and to prevent any loss of evidence which could occur. Self-inflating tents can also be used as a body-holding area before removal to the mortuary, or as a work station for the forensic recovery team. If there is any doubt as to whether this equipment is contaminant-free, new equipment will be called for.

As most investigations at a bomb scene last some time, at least part of the investigation will take place during the night. To facilitate this, lighting with its own power source can be used. However, in the event of rain or wet ground, particularly on tarmac surfaces, light reflection can lead to vital evidence being missed by the naked eye. In such cases it may be necessary for the bomb-scene manager to suspend that part of the recovery and deploy the forensic recovery team on another aspect of the investigation, rather than risk losing evidence. It may even be prudent to stop the forensic recovery completely until daylight permits further progress. Another factor to be taken into account by the bomb-scene manager is that, while lighting may be sufficient to enable work to be carried out, other areas will be placed in deep shadow, which may increase the nature of potential hazards due to the lack of visibility outside the work area.

The examination of each zoned area begins with a walk through by the forensic recovery team, walking in line abreast to see if any parts of the device are visible to the naked eye and can quickly be retrieved. Behind the line is a nominated officer, to whom any item is passed; the item is then sealed in the appropriate bag and identified. Having completed the 'walk through', the area is then painstakingly swept with brooms. All items of debris and loose material are swept into piles; then, time and the nature of the bomb dictating, the debris is either sifted at the scene by hand, in order to single out component parts of the device, or shovelled into bags, which are then sealed and placed in boxes for a more detailed examination later. All the equipment used by the officers during this process is either brand-new or has been sterilized before use to avoid any suggestion of contamination. Each zone is dealt with in a similar fashion, except for the seat of the explosion itself. Any vehicles within a zone are inspected in detail for fragments of the device or items of evidential value. Once the area surrounding a vehicle has been cleared and the vehicle examined, it is then moved to a cleared area so that the ground shielded by the vehicle may be searched. The removal of vehicles can present their own difficulties, particularly if large and locked. In such cases, cleared, taped paths are created to allow access for removal vehicles. The decision as to when and how to use such equipment is only made after taking into account contamination issues and the possible loss of evidence.

As each zone is cleared, the local authorities, after consultation, can be allowed to move in to begin the clearing-up process. This only takes place if the geography of the scene allows them a controlled entrance and exit with no possible overlap to the areas not yet examined. The contamination issue would be considered in detail.

This process takes considerable time, which, depending on the scene and the spread of damage, can last many days. One of the essential duties of the

bomb-scene manager is therefore regularly to update the senior investigating officer and the local police commander, to facilitate contact with the media and to inform the community and other interested agencies.

The forensic recovery team must be supplied with sufficient food and drink during what may be a lengthy investigation. To avoid loss of manpower and time-wasting, the food is brought to a convenient area within the cordons, preferably adjacent to the control van. Hot and cold drinks are particularly essential.

The zone requiring the most painstaking search is the actual area of the bomb site itself. The area is swabbed and the swabs are despatched to the forensic explosives laboratory so that scientists can determine the nature of the explosives substance and advise the investigating officers accordingly. Under the direction and control of the exhibits officer, the forensic recovery team, on hands and knees, conduct a fingertip search of the entire area. The area is then swept and the debris collected. The crater at the site of the explosion is dealt with by the exhibits officer in a step-by-step process. This involves measuring the crater and then removing all loose debris, which is treated as a single exhibit. The crater is then once again measured and then the entire area is excavated normally to at least 0.3 m, depending on the ground. This again is treated as a separate exhibit. The entire process is photographed during each phase.

Any vehicles that may be of evidential value to the forensic scientists or supply clues for the investigating officers (normally vehicles in the immediate vicinity of the bomb or vehicles that had the bomb inside or underneath) are removed for forensic examination. The area surrounding the vehicles is cleared, as previously explained, and the vehicle is then lifted by mechanical means and placed on a large tarpaulin. A second tarpaulin is placed over the top and the entire vehicle is parcel wrapped. The vehicle is then loaded onto a transporter and taken to the forensic explosive laboratory for a detailed examination. This method of packaging ensures that any evidence from the vehicle is contained within the tarpaulins and no outside contaminants can be added. A nominated officer will accompany the vehicle to the laboratory and hand it over to a liaison officer to ensure continuity of the evidence trail, which is the case for all items submitted for examination.

The entire process is meticulously recorded by the exhibits officer from his or her arrival on the scene until the forensic team's withdrawal. Every item is bagged and numbered and each stage of the process carefully logged and noted. It is these notes which the exhibits officer will later rely on, should he or she be required to give evidence in court. Throughout the forensic retrieval process liaison is maintained with the Forensic Explosive Laboratory, seeking advice according to circumstances. If necessary, scientists can be asked to attend the scene and advise as to evidence retrieval and best practice. This assistance can be of vital importance in determining the correct method of retrieval and submission for examination.

No bomb scenes are the same; they can vary substantially in size and impact. Each will have unique difficulties for the bomb-scene manager and team to deal with. In some cases it may be necessary to summon fingerprint experts to examine areas that, if disturbed, may lead to the loss of evidence. Once again, protective clothing will be provided. In many cases serious damage may have been sustained by the various utilities, causing large areas of disruption to the community and a serious hazard within the scene. Experts will be called to repair or isolate such damage and their operations must be facilitated in order to provide a return to normality as soon as possible. The scene of some explosions may encompass areas of high priority or sensitivity, which once again may produce areas of difficulty that must be resolved. Consideration will also be given, in certain conditions, to call for an air exclusion zone, as the down-draft from helicopters can prove extremely dangerous to those on the ground. In the event of rain, it may be necessary to seal drains to prevent evidence being washed away, and, again under certain circumstances, the local authority may be called upon to open drains, culverts, etc. so that they may be fully examined.

When the forensic retrieval is complete the entire area is once again photgraphed from numerous angles and vantage points. The scene is then handed back to the local police commander.

See also: **Causes of Death**: Overview; Scene of Death. **Crime-scene Investigation and Examination**: Recording; Collection and Chain of Evidence; Recovery of Human Remains; Packaging; Preservation; Contamination; Major Incident Scene Management. **Explosives**: Analysis; Mechanism of Explosion.

Further Reading

Brodie TG (1996) *Bombs and Bombings: A Handbook to Detection, Disposal, and Investigation for Police and Fire Departments*. Springfield, IL: Thomas.

Committee on Smokeless and Black Powder, Board on Chemical Sciences and Technology, Commission on Physical Sciences, Mathematics, and Applications (1998) *Black and Smokeless Powders: Technologies for Finding Bombs and the Bomb Makers*. Washington, DC: National Academy Press.

Cundill JP (1889) *A Dictionary of Explosives*. Chatham: Mackay.

Marshall A (1917–1932) *Explosives*. Philadelphia: Blakiston.

Powell W (1971) *The Anarchist Cookbook*. New York: Stuart.

Urbanski T (1965) *Chemistry and Technology of Explosives*. Oxford: Pergamon Press.

US Bureau of Alcohol, Tobacco, and Firearms (1979) *Explosives Tagging: Scientific Finger Printing for Law Enforcement*. Washington, DC: GPO.

US Bureau of Alcohol, Tobacco, and Firearms (1990) *ATF – Explosives Law and Regulations*. Washington, DC: GPO.

US Congress House Committee on Science, Space and Technology (1989) *Airport Security Technology and Research Act of 1989: Report (to accompany HR 2365)*. Washington, D.C: GPO.

US Congress Office of Technology Assessment (1980) *Taggants in Explosives*. Washington, DC: GPO.

US General Accounting Office (1986) *Ammunition and Explosives: Improved Controls and Accountability at Fort Bragg: Briefing Report to the Honorable Pete Wilson, US Senate*. Washington, DC: GPO.

Yinon J and Zitrin S (1981) *The Analysis of Explosives*. Oxford: Pergamon Press.

Yinon J (1999) *Forensic and Environmental Detection of Explosives*. Chichester: Wiley.

Commercial

S G Murray, Cranfield University, Royal Military College of Science, Shrivenham, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0508

Introduction

To secure the best value from this article it is recommended that the article entitled 'Mechanism of Explosion' be read first.

The first use of explosives for commercial enterprises dates back to the seventeenth century, for example for mineral mining in Hungary, tin mining in Cornwall in the UK and civil engineering in France with the building of the Languedoc tunnel. In all of these cases the explosive used was black powder. Even today black powder still finds limited commercial use for example in slate quarrying where the shattering effect of a high explosive would be too damaging to the fragile slate.

The true potential of explosives for mining and quarrying began with the development of nitroglycerine-based explosives in the nineteenth century and the advent of the detonator. This created a much more efficient method of using chemical energy both in breaking rock and for lift and heave in trenching or cratering to move earth. These dynamites or gelignites were the mainstay of commercial explosives until the 1950s at which time explosives based on ammonium nitrate began to develop. This has led to the development of slurry and emulsion explosives and these are now displacing the nitroglycerine explosives in the commercial sector.

In comparison to military high explosives, commercial explosives are driven by cost. This is because the explosive is a significant component in the costs of a commercial enterprise such as quarrying or tunneling. It may be more cost effective, particularly in recent times, to use mechanical methods instead of explosives. The Channel Tunnel between the UK and France was excavated without the use of any explosives. Furthermore, in relation to costs, as the shattering ability or brisance of the explosive increases, so does the cost of the explosive. This means that there is a need for a range of explosives with different properties, particularly their detonation pressure but also their power, that is gas expansion effects.

Performance Parameters

Indication of performance

For commercial explosives the ability to create lift and heave (power) is usually described by strength. This is a comparison of a particular explosive with a standard. The comparison may be with the same weight (mass) or the same volume of the standard giving the 'weight strength' or 'bulk strength', respectively. For nitroglycerine-based explosives, strength is usually compared to blasting gelatine, the most powerful of the type. More recently, strength may be compared to other standards and thus the literature must be scrutinized carefully to assess the actual performance. The method of conveying the information either in the literature or on the wrapping or container of the explosive charge is to quote the strength as a percentage, e.g. 80% strength.

The manufacturer rarely gives the detonation pressure; however, the literature will give the density and the detonation velocity. Together, these will indicate the shattering ability as high velocity of detonation and high density (relative) give high detonation pressure. This is indicated in the description of the various types of explosive in the following text.

Fume characteristics

One of the hazards of using explosives in confined spaces such as tunneling is the production of toxic fumes. Commercial explosives are mixtures of various ingredients and it is possible to vary the ratio of fuel elements to available oxygen. This means that the manufacturer can create explosive formulations that minimize the production of carbon monoxide thus making them much safer to use in confined situations.

For some explosives it is possible to create toxic products by having an excess of oxygen relative to the fuel that is present. This is particularly true for ammonium nitrate/fuel oil where a deficiency of fuel may lead to the production of oxides of nitrogen, seen as a brown gas cloud after detonation.

Water resistance

Another property of commercial explosives that is important to the user is the ability to use the explosive in the presence of water, for example in wet boreholes. Thus, the details provided for a particular product would state the water resistance from 'none' to 'excellent'. This water resistance may be due to the packaging in more recent times with the advent of plastic containers. However, in earlier times it was a property of the actual formulation. For example, ammonium nitrate/fuel oil as a loose material has no water resistance, whereas the gelignites with a high nitroglycerine content have excellent water resistance.

Shock sensitivity

An important factor in the use of commercial explosives is the knowledge of the ability to reliably detonate the explosive charge with a detonator. This is known as detonator or cap sensitive. The formulations containing nitroglycerine are detonator sensitive; however, many of the bulk borehole charges are not. These include ammonium nitrate/fuel oil and certain slurries and emulsions. However, it is possible to formulate slurries and emulsions that are detonator sensitive.

For charges that are not detonator sensitive, a booster (or primer) is employed. This is an explosive charge that will reliably detonate from a detonator and amplify the shock wave to then detonate the insensitive column of explosive.

Critical diameter

Critical diameter is the minimum diameter of a bare charge of explosive that will sustain a detonation shock wave. For military explosives this value is very small, often 1–2 mm. However, for commercial explosives the value can be much greater, perhaps as much as 50 mm. This requires a matching of explosive type to borehole diameter. In practice, the manufacturer will supply cartridged explosives in diameters that are guaranteed to detonate fully. If, however, the explosive is to be bulk loaded into a borehole it is imperative that the hole diameter is well above the critical diameter for the explosive.

Nitroglycerine-containing Explosives

From the mid-nineteenth century to the mid-twentieth century, nitroglycerine (NG) (**Fig. 1**) was the most important energetic ingredient and sensitizer for commercial explosives. The early developments were attributed to the Nobel family in Sweden with Immanuel Nobel being the first to realize the potential of nitroglycerine and then his son Alfred who has now been immortalized through his efforts in developing modern explosives and the introduction of the Nobel Prizes. The oily liquid is manufactured by the reaction of glycerine (glycerol, propane-1,2,3-triol) with a mixture of concentrated nitric and sulfuric acids, during which the temperature must be controlled carefully to avoid a rise in temperature that could cause a runaway decomposition of the explosive that has formed resulting in an explosion. Modern synthesis uses a continuous nitration process where the maximum quantity of explosive in the reaction vessel is limited.

Pure nitroglycerine has a melting point of 13°C and it is undesirable for the explosive to freeze when mixed in a formulation. Partly thawed nitroglycerine is dangerously sensitive; this is thought to be due to the presence of triclinic crystals that may rub together if the charge is handled. To overcome this, another explosive molecule, ethyleneglycoldinitrate (EGDN, nitroglycol) (Fig. 1) is mixed with the nitroglycerine to lower the freezing point. EGDN is manufactured at the same time as the nitroglycerine by using a mixture of glycerine and glycol (ethane-1,2-diol) in the nitration reaction. The ratio of these explosive types is usually around 50/50 and lowers the freezing point of the mixed liquid explosives to around −10°C. Pure nitroglycol freezes at −22°C.

The range of explosives described below is manufactured worldwide; however, in recent years the development of commercial explosives that do not

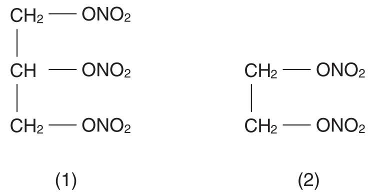


Figure 1 The structure of nitroglycerine (1) and ethyleneglycoldinitrate (2).

contain nitroglycerine is causing a serious decline in the use and therefore production of this type of explosive. It is estimated that in the very early part of the twenty-first century, the production of dynamites will cease.

Dynamite explosives

This group of explosives was developed from the original dynamite of Nobel that was simply nitroglycerine absorbed into kieselguhr, a dry powdered clay. This significantly reduced the extreme sensitivity of the nitroglycerine to shock initiation. There were two major developments that followed: (1) the addition of nitrocellulose (NC) that gelled with the nitroglycerine (and nitroglycol as mentioned above); (2) the inclusion of a fuel/oxidizer mixture in the formulation. The use of nitrocellulose allowed the liquid nitroglycerine/nitroglycol to form a gelatinous mixture giving useful physical properties and minimizing the separation of the liquid explosives from the mixture. Also, it significantly reduced the sensitivity of the pure liquid explosives. This led to a range of explosives known as gelatine dynamites or gelignites with the NC content being approximately in the ratio 1:20 with the NG/EGDN. The original dynamite as described above is no longer found.

The use of a fuel/oxidizer mixture gave the extra dynamites or extra gelatine dynamites although it should be noted that these are often simply called gelignites. The fuels are usually cellulose based, such as sawdust or wood meal (fine sawdust), and the oxidizer is either sodium or, more usually, ammonium nitrate. In some countries the term 'straight gelatine dynamite' may be used to differentiate the use of sodium rather than ammonium nitrate and 'ammon' or 'special' to denote the use of ammonium nitrate.

This allows a large range of explosive formulations to be produced depending on the ratio of gelled nitroglycerine/nitroglycol to fuel/oxidizer mixture. Those with a reasonable gelled nitroglycerine/nitroglycol content are rigid in form and are the true gelatine explosives. They can be deformed by pressing with a tamping rod when loading shot holes. As this ingredient is reduced in the formulation then the consistency becomes more crumbly. These are known as semigelatines and will have reduced performance. Finally, the content may be reduced so much that the consistency is a powder; at this stage it is unlikely that nitrocellulose is present. These low performance explosives are the nitroglycerine powders. In some countries it is possible to find formulations that have the same NG/EGDN content as a gelatine or semigelatine explosive but without the NC as a gelling agent. Confusion may occur as these have been called 'straight' dynamites in some texts.

Other additives may be found in these formulations. Both barium sulphate and manganese dioxide can be added to increase the density and achieve maximum velocity of detonation. Sodium chloride and calcium oxalate are added to act as flame suppressants (see below) and often naturally occurring gums are present to aid consolidation of the mixtures. Calcium carbonate may be added to act as a stabilizer to minimize the risk of autocatalytic decomposition of the nitroglycerine/nitroglycol. On rare occasions, dinitrotoluene may be present.

Permitted explosives

When blasting in coalmines there is always the risk of secondary explosions and fire from ignition of either methane/air or suspended coal dust/air mixtures. To overcome this problem permitted explosives were developed. These must pass rigorous national testing to insure that such ignitions cannot occur and are categorized for particular uses. They are given designations such as P1–P4/5 in the UK.

The source of ignition for the fuel/air mixtures that may be present could be one of the following:

1. A long lasting flame as would be produced from a black powder explosion;
2. A secondary burn of hydrogen/carbon monoxide produced by an explosive that had a negative oxygen balance;
3. An intense shock wave from a high detonation pressure explosive causing shock heating of the surrounding air.

To overcome these possible mechanisms formulations have been devised that have relatively low shock pressure and a positive oxygen balance. Furthermore, sodium chloride is added in significant quantity, as much as 30% by weight, as this acts as a flame suppressant by interfering with the flame propagation process.

Performance

Nitroglycerine is a very powerful explosive producing an energy release of $6275\,J\,g^{-1}$ and $740\,cm^3\,g^{-1}$ of gas. This ranks with the highest performing explosive molecules for power output. Nitroglycol has an even higher energy output at $6730\,J\,g^{-1}$ but the same gas production. It is clear that a high percentage of these ingredients will produce an explosive with good lift and heave. Also, the velocities of detonation and densities suggest a high detonation pressure. However, as discussed above it is usual to find these explosive liquids absorbed into a nitrate/wood meal mixture and this will significantly reduce the density and velocity of detonation.

The most energetic formulation of this group of explosives is blasting gelatine, which contains 92–94% nitroglycerine/nitroglycol together with 6–8% nitrocellulose. This is used as the standard against which the strength of other gelatine dynamites is measured. This explosive is rarely used in practise, as the level of performance is hardly ever required. As the percentage of nitroglycerine/nitroglycol decreases the detonation pressure falls quite rapidly. However, although the power also falls the percentage reduction is not as great as for the detonation pressure.

Typical percentages of nitroglycerine/nitroglycol for the various types described above will be:

Blasting gelatine	92–94%
Gelatine dynamites	75–25%
Semigelatines	20%
NG powders	10%
Permitted	30–10% (being gelatines, semigelatines or NG powders)

The weight strength, bulk strength, density and velocity of detonation for typical examples of the various types are given in **Table 1**. The velocity of detonation is the maximum as achieved by a strong shock initiation. These types of explosives have an unusual feature if initiated with a relatively weak shock. They will detonate with a velocity of detonation of $\sim 2500\,m\,s^{-1}$ irrespective of the NG/EGDN content. Thus a gelatine, semigelatine or NG powder can have two quite different values for a stable velocity of detonation.

Table 1 also shows that the bulk strength often is lower than the weight strength. This is because in those cases the density of the explosive is much lower than the blasting gelatine standard.

Military Dynamites

It should be noted that in the United States, military dynamites are not based on nitroglycerine but contain trinitrotoluene (TNT) and cyclotrimethylene trinitramine (RDX) as the explosive molecular components.

Ammonium Nitrate-based Explosives

Ammonium nitrate fuel oil (ANFO)

Any carbonaceous material mixed with ammonium nitrate (AN) will produce a possible explosive mixture. In fact AN as a pure material is not classified as an explosive for transportation and storage unless it contains $>0.4\%$ carbonaceous material. However, there have been many apparent incidences of pure AN being transported in bulk that has caught fire and ultimately transitioned to a detonation. In all cases where an explosion followed an AN fire, the mass of AN was confined, for example in the hold of a ship, and it is contested was almost certainly contaminated with combustible materials to some extent. The infamous incident at Oppau, in the Rhineland of Germany, was caused by the explosive blasting of an approximately 50/50 mixture of AN and ammonium sulfate. Although it was found that there had been around 20 000 blasts carried out previously, a procedure used to break up the hard consolidated mass of AN/ammonium sulfate, in this instance it is likely that the condition of the material had altered from the norm and probably contained less moisture and was of slightly lower density. The ensuing explosion killed 500, injured 1900 and caused damage in a town 1.5 km distance. Most of Oppau was completely destroyed. The explosion was estimated to be equivalent to around 500 tonnes of TNT.

Ammonium nitrate (NH_4NO_3) has been manufactured as a fertilizer at least since the turn of the twentieth century. Developments in the 1950s led to a form of AN known as prills. These were small spherical or ovoid beads, 2–3 mm in diameter, manufactured by passing concentrated AN solution droplets down a drying tower against a rising stream of hot air. Control of the conditions gave prills of differing densities. The crystal density of AN is $1.7\,g\,cm^{-3}$. For agricultural use a high density prill is favorable at $\sim 1.1\,g\,cm^{-3}$; however, for use in the manufacture of an explosive a lower density is better at $0.8–0.9\,g\,cm^{-3}$. This is because the porosity of the lower density material allows the liquid fuel (fuel oil

Table 1 Performance figures for nitroglycerine-based explosives

Explosive type	*Weight strength vs BG (%)*	*Bulk strength vs BG (%)*	*Density ($g\,cm^{-3}$)*	*Velocity of detonation (ms^{-1})*
Blasting gelatine	100	100	1.6	7500
Gelatine dynamite	80–65	80–65	1.5–1.4	6600–5500
Semigelatine	65–50	60–50	1.3	6000–5000
NG powder	80	65	1.15	4000

or diesel) to absorb readily into the prills. Thus, an intimate fuel/oxidizer mixture is achieved. These lower density prills are somewhat prone to physical breakage, which at first appeared to be a disadvantage. However, with some modern methods of use it may actually be advantageous as described below.

The commercial explosive is known as ANFO (AN fuel oil) and the correct proportions of AN to diesel, a commonly used fuel oil, are 94/6 based on pure AN. If it is known that the AN is not pure then the diesel percentage should be adjusted to give the same ratio of the AN content to the diesel. It is important to ensure that the fuel/oxidizer balance is correct as variance may give large quantities of carbon monoxide if fuel rich, or of nitrogen dioxide if oxidizer rich. Most explosives manufacturers will sell both ready-made ANFO and bags of AN prills for the user to manufacture the ANFO on site. For large-scale use it is mixed on site in a purpose-built vehicle and poured directly into vertical boreholes. ANFO can be loaded into horizontal or vertically inclined boreholes by pneumatic loading. This is carried out at a pressure of $\sim$4 bar which causes the density to rise from the 0.8–0.9 g cm^{-3} to $\sim$1.1 g cm^{-3}. This provides an explosive fill that has more power and a higher velocity of detonation.

Performance of ANFO

ANFO is a very insensitive explosive, which makes it extremely safe to use. However, it is not detonator sensitive and therefore requires the use of a booster to provide reliable detonation. When poured loose into a borehole or pneumatically loaded, there will be maximum filling of the hole. This does not occur when cartridged charges are used.

On the other hand, ANFO has no water resistance unless packaged in waterproof containers. It cannot be loaded directly in to wet boreholes as this makes the column of explosive insensitive even to a booster. Also, ANFO has a large critical diameter and is not likely to propagate a detonation shock wave successfully in boreholes of less than 50 mm in diameter. **Figure 2** indicates how the velocity of detonation varies with borehole diameter. As the value increases so will the shock pressure. A commonly used borehole size is 100 mm (4 inches).

Typical values quoted for the weight strength and bulk strength of ANFO are 70% BG (blasting gelatine) and 30% BG, respectively. The bulk strength is so low because of the very low density of poured ANFO. Addition of aluminum powder increases the energy in the mixture and gives weight and bulk strengths of 77% BG and 40% BG, respectively. The shock pressure of these explosives will be in the

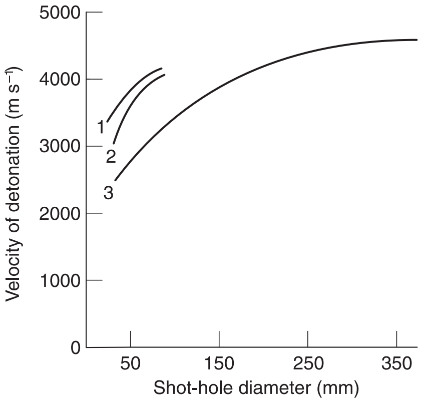


Figure 2 Velocity of detonation for ANFO (poured and pneumatically loaded) versus borehole diameter. 1, Pneumatically loaded, 1.1 g cm^{-3}; 2, pneumatically loaded, 0.95 g cm^{-3}; 3, poured ANFO, 0.8 g cm^{-3}.

region of 10–30 kbar. The higher values will be for large diameter boreholes as the velocity of detonation is higher or for the pneumatically loaded ANFO where the density is higher. A typical value for velocity of detonation for poured ANFO in a standard borehole of 100 mm would be $<$3000 m s^{-1}.

Slurry explosives

A slurry explosive is based on AN as oxidizer mixed with a fuel as with ANFO. However, they have as part of the formulation a significant percentage of water, usually 10–25% by weight. The water, together with the presence of guar gum or equivalent as a gelling agent, produces an explosive superior in many ways to ANFO. The basic formulation is ammonium nitrate solid together with either a carbonaceous fuel or aluminum powder or both held in suspension in a gelled saturated solution of ammonium nitrate. Over the years many types of carbonaceous fuels have been used, ranging from solid hydrocarbons, powdered coal, carbohydrates (e.g. sugars), bagasse (sugar cane cellulose) to paint-grade aluminum. When aluminum is present in the mixture it is usual to include an additive to buffer the pH against the acidity caused by the presence of the ammonium nitrate. The pH is held at about 5 to prevent reaction with the aluminum that would evolve hydrogen.

To provide sensitivity to shock initiation it is usual to introduce very small (40–50 μm) air pockets or bubbles, which act as centers that create hot spot initiation when a shock wave passes. This is done in one of several ways the simplest of these being to use a beating process and the air bubbles are trapped as the gel sets. Other methods are to add glass or plastic microballoons or expanded polystyrene beads. These

are favored if it is thought that the pressure in a borehole may squeeze the explosive increasing the density and reducing the number of air pockets or if the shock pressure from an adjacent borehole creates a similar effect. A recent chemical reaction method has been used to create the bubbles. This is the reaction between sodium nitrite and acetic acid which produces nitrogen gas bubbles. The presence of aluminum powder also leads to hot spot creation and in some cases actual explosive molecular types are added such as smokeless powder (propellant) or 20-mesh TNT. The most recent developments have led to the use of nitrated chemical sensitizers that introduce both sensitivity and fuel, examples being isopropyl nitrate or the now favored methylamine nitrate.

Until relatively recently, the manufacture of these complex mixtures could be achieved only under factory conditions. Thus, slurry explosives were supplied cartridged in flexible plastic tubing similar to that used to package meat products such as liver sausage. Current cartridge sizes have diameters of 25–200 mm and weigh between 200 g and 5 kg. In more recent times systems have become available for on-site manufacture and mixing trucks are available that manufacture and pump the product directly into the borehole. A further development has been the use of blends of slurry with ANFO (see below).

Performance of slurry explosives

At first, slurry explosives were detonator insensitive and required a booster as for ANFO. Developments have now led to compositions that are detonator sensitive, but the majority still require boosting. Due to their physical nature as water-containing gelled systems they are quite insensitive to accidental initiation. The use of gelling agent gives these explosives a reasonable degree of water resistance if loaded unpackaged into a wet borehole and thus is a superior explosive to ANFO. Most slurry compositions have reasonably large critical diameters and are employed in boreholes greater than 75 mm in diameter although as noted above, the smallest size is a 25 mm diameter cartridge.

As explosives they are superior to ANFO due to a large extent to the higher densities that are achieved. Densities are in the range of 0.9–1.4 $g\,cm^{-3}$ and this, together with the higher energies gives velocities of detonation in the range 3500–5500 $m\ s^{-1}$ and this will give shock pressures in the range 50–100 kbar. Depending on the formulation, weight strengths are in the range of 50–100% BG and bulk strengths from 30–75% BG. The much higher density than ANFO produces bulk strengths that are relatively not as low compared to the weight strength as is seen for ANFO.

Emulsion explosives

At first sight the formulation of an emulsion explosive appears very similar to a slurry being based on ammonium nitrate, water, hydrocarbon oil, aluminum powder and a sensitizer. However, there is a fundamental difference in the physical nature of the two explosives. A slurry has a continuous aqueous phase, whereas the emulsion consists of aqueous droplets containing the ammonium nitrate held in a true emulsion with a hydrocarbon continuous phase. The hydrocarbon layer may be as little as 3 μm in thickness. Depending on the properties of the hydrocarbon, from a mobile oil to a wax at ambient temperatures, the consistency of an emulsion explosive ranges from toothpaste to putty.

Other than this, there are a number of similarities between emulsions and slurries. The sensitization is achieved by exactly the same methods as described above. Both factory cartridged and truck-mixed versions are used and the critical diameter considerations are about the same although those with the higher densities are detonator sensitive and maintain the detonation in small diameter, e.g. 25 mm.

A current technique in the use of emulsion explosives is to use a mixture with ANFO. This fills the voids in the standard ANFO fill creating not only a denser material but also one of higher energy. The mixtures range from 70/30 emulsion/ANFO to 30/70 emulsion/ANFO. The former can be pumped but the latter is like thick porridge and is augered into the boreholes.

Performance of emulsion explosives and emulsion/ANFO blends

The range of emulsion explosives is very similar to the slurries in explosive performance for velocity of detonation, weight strength and bulk strength. However, advantage is gained by using the blend technique. **Table 2** indicates the change in density and energy between ANFO, an emulsion and a typical blend. The velocity of detonation tends to be higher for the blend than either component and can approach 6000 $m\ s^{-1}$ at a density of $\sim$1.4 $g\,cm^{-3}$.

Miscellaneous types

There are two types of ammonium nitrate-based explosives that are available but are peripheral to the mainstream commercial production of explosives. In one the AN is mixed with the liquid nitromethane (CH_3NO_2), perhaps up to 15% of the liquid to produce an explosive with a velocity of detonation of $\sim$5500 $m\ s^{-1}$ and shock pressure of $\sim$90 kbar. In the other the AN in mixed with hydrazine (N_2H_4) to produce a trade product known as ASTRO-PAK or

Table 2 Performance comparison between ANFO, an emulsion and an ANFO/emulsion blend

Property	*ANFO*	*Emulsion*	*ANFO/emulsion blend*
Density (g cm^{-3})	0.8	1.25	1.3
Energy (kJ g^{-1})	3.8	2.9	3.6
Energy (kJ cm^{-3})	3.1	3.6	4.5
Velocity of detonation (m s^{-1})	3000	4500	5500

Astrolite T. The explosive is a liquid with a density of ~1.4 g cm^{-3} and a velocity of detonation of 8000 m s^{-1} if initiated by a strong shock. This will give a detonation shock pressure of >200 kbar and thus the ability to break metal.

Detonating Cords

Essentially, detonating cord is a plastic tube, sometimes reinforced with wound or woven fibers and filled with powdered explosive. It is used to transmit a detonation shock wave for multiple charge initiation, that is to link charges together, or to lead a detonation shock wave from a detonator to a booster that may be in a borehole. Also, it may be used as an explosive charge in its own right.

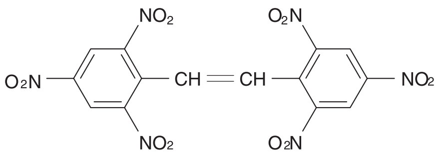
There are three parameters relevant to the description of a particular cord. One is the tensile strength, which is important if any load may be applied to the cord. This tends to be relevant only for use in oil wells. The second is the type of explosive used to fill the cord and the third the loading of explosive per unit length. Obviously, as the loading increases so will the diameter of the cord. The smallest commercial cord contains 1 g m^{-1} of explosive and the largest 100 g m^{-1}.

In general use, cords containing 5–8 g m^{-1} are used to lead a detonation to a booster or to initiate shock tube (see below). The cords used for linking multiple charges contain 10–12 g m^{-1} and the cords with 40 g m^{-1} upward are used for engineering operations such as presplitting of rock or for demolitions.

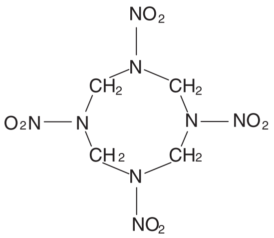
The general explosive filling is pentaerythrytol tetranitrate (PETN) as this explosive has a small critical diameter of ~1 mm and thus will propagate a detonation shock wave reliably in very small columns. However, when cords are used in oil wells there may be a problem associated with the temperature in the deeper wells. In down-hole operations it is necessary to have the explosive charges in the hole for some time before firing. The cords filled with PETN can be used at a maximum temperature of 135°C for 1 h or 121°C for 24 h. For higher temperatures different explosives are used, structures of which are given in **Fig. 3** and approximate maximum operating temperatures in **Table 3**.

Boosters (Primers)

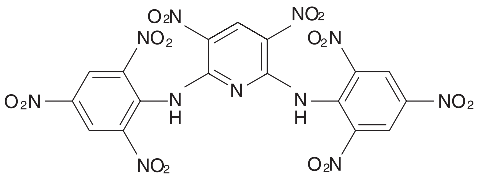
As discussed above, many commercial explosives are not detonator sensitive. Reliable initiation is achieved



(1)



(2)



(3)

Figure 3 Structures of some explosives used at high temperatures. 1, hexanitrostilbene; 2, cyclotetramethylenetetranitramine; 3, 3,5-dinitro-2,4-di(picrylamino)pyridine.

Table 3 Approximate maximum operating temperatures for prolonged residence times

Explosive	*Maximum operating temperature (°C)*
Pentaerythrytoltetranitrate (PETN)	125
Cyclotrimethylenetrinitramine (RDX)	165
Cyclotetramethylenetetranitramine (HMX)	180
Hezanitrostilbene (HNS)	300
3,5-dinitro-2,4-di(picrylamino)pyridine (PYX)	>300

through the use of booster charges that provide an intense shock wave to the column of explosive in the borehole. The most commonly used explosive in a booster is Pentolite, a 50/50 mixture of PETN and TNT. This has a detonation shock pressure of ~250 kbar and reliably detonates ANFO, slurries and emulsions.

Detonators

Standard demolition detonators

For commercial use it is normal for electric detonators to have a time delay from the firing current initiating the fuze head to the detonator providing the shock output. This is done by introducing a delay element consisting of a column of pyrotechnic between the fuze head and the lead azide pellet. The concept of delay detonators is to allow one firing signal to be sent to multiple detonators that will initiate all the fuze heads; however, for different delay times this will give different firing times for the detonators. This reduces ground shock when quarrying by separating borehole detonations by intervals of ~25 ms. In tunneling it allows some charges to fire first to create a face towards which the next row of charges firing some time later can work. Finally, for demolitions it is necessary for some parts of a structure to be blasted first giving a controlled collapse or a direction for toppling. There are two delay series; the millisecond series separated by 25 ms intervals up to 850 ms for some manufacturers and the half-second series usually up to a maximum of 6 s.

Delay detonators can be identified and are classified by the delay number stamped on the end of the detonator body or on a plastic or paper tag attached to the leg wires. This is usually a number that must be related to the literature of the manufacturer to ascertain the delay time. This literature will describe the color of the leg wires and the type of wire (iron, copper, etc) and the material for the body, normally copper or aluminum.

There are nonelectric detonators known as 'plain' detonators. These are essentially an electric detonator without the fuze head and associated leg wires. To initiate this a delay fuze is inserted into the open end of the tube and crimped in place with a special tool immediately before use. The delay fuze usually contains black powder and burns at a steady rate of ~300 mm in 40 s. The flash output from the fuze initiates the lead azide pellet and fires the detonator. These are commonly used for single shots such as breaking a large rock on the quarry floor or in tree stump removal.

Special initiating systems

There have been several examples of developments aimed at improving the safe use of detonators. The Magnadet detonator was developed in the UK by ICI Explosives Ltd as a system that was immune from induced current initiation caused by electromagnetic radiation. This problem exists for fuse head-containing detonators, as the majority will fire from a current as low as 0.5 amp. This can be induced if an electric detonator is used near a powerful radio, radar or under electric power cables. The Magnadet is a normal electric detonator with very short leg wires which form a continuous circuit with several loops wrapped around a ferrite toroid. The firing cable is not attached electrically to this circuit but passes through the hole in the toroid. When a special firing signal is sent down the firing cable it induces a current in the detonator circuit and fires the detonator.

Perhaps the most important development in recent years has been the shock tube detonator. This system relies on the initiating signal being transmitted to a detonator via a small hollow plastic tube ~2 mm in diameter the inside of which is coated with a mixture of the explosive cyclotetramethylenetetranitramine (HMX) and aluminum powder at a loading of ~20 mg m^{-1}. This layer propagates a shock wave at ~2000 m s^{-1} which on reaching the detonator initiates the delay element if present or the lead azide pellet directly. There is so little explosive present that the tube is not disrupted and indeed could be held as the shock wave passes down the tube. The shock tube is factory fitted to the detonator and is available in various lengths. Systems are available to link tubes to

create bunch firing. The advantage of this system is that it is completely immune to initiation from stray electric currents or induced current from electromagnetic radiation.

A recent development has been the introduction to the market of detonators that do not contain lead azide. It has been replaced with a finely powdered secondary explosive, PETN, which in a confined metal collar burns to detonation. The advantage of using a secondary explosive instead of the primary explosive, lead azide, is that the detonator is much more difficult to initiate accidentally. Trials in which rocks and weights were dropped on to both types of detonator demonstrated this insensitivity.

See also: **Explosives**: Analysis; Mechanism of Explosion; Military; Bomb-scene Management.

Further Reading

Atlas Powder Co. (1987) *Explosives and Rock Blasting, Field Technical Operations*. Dallas: Atlas Powder Co.

Dolan JE and Langer SS (1997) *Explosives in the Service of Man, the Nobel Heritage*. Cambridge: Royal Society of Chemistry.

Gregory CE (1993) *Explosives for Engineers*, 4th edn. Zurich: Clausthal-Zellerfeld.

Hopler RB (1998) *Blasters Handbook*, 17th edn. Cleveland: International Society of Explosives Engineers.

Olofsson SO (1991) *Applied Explosives Technology for Construction and Mining*. Arla, Sweden: Applex.

Persson P-A, Holmberg R and Lee J (1994) *Rock Blasting and Explosives Engineering*. Boca Raton: CRC Press.

Wang X (1994) *Emulsion Explosives*. Beijing: Metallurgical Industry Press.

Mechanism of Explosion

S G Murray, Cranfield University, Royal Military College of Science, Shrivenham, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0506

Introduction

An explosive can be defined generally as a store of energy that can be made to release that energy in such a manner as to do work. This energy release may be in a burning reaction as for a gun or rocket propellant or in a detonation as for a military warhead or a commercial blasting explosive. Explosives that burn in their designed use sometimes are called 'low explosives', however, in virtually all cases these are propellants of one sort or another. Those that detonate in their designed use are known as 'high explosives'. A further type of explosive is the pyrotechnic. These are also self-contained energy stores although the energy release is usually not designed to do work, but to produce a special effect such as light (white or colored), sound, smoke, delays, etc. and as such is on a longer timescale than for propellants or high explosives. The one thing that is common to all these systems is that they all create their effects through a chemical reaction.

Chemical Source of Energy for an Explosive

The combustion process is the universal method of producing heat for industrial or domestic purposes. A fuel consisting of materials that contain carbon and hydrogen is allowed to react with a source of oxygen, which in this case is the air. The combustion of 1 kg of hydrocarbon, such as gasoline, releases around 45 MJ converting the fuel into carbon dioxide and water (CO_2 and H_2O). When used for industrial or domestic purposes this reaction proceeds in a controlled manner such that the energy release and gas production is relatively slow. However, we know to our cost that if a volatile fuel is released into the air in an uncontrolled manner, mixes with air and then is ignited, a fuel–air explosion occurs. This has been the cause of many accidental explosions worldwide.

Thus, for this combustion process to directly cause damage, or do work, it is not just the amount of energy that is released, but the speed at which that release occurs. Also, to do work requires an agent that can transfer energy. In the case of an explosion this is the gases that are formed during the combustion. Thus, for propellants and high explosives it is necessary to find a system that can produce heat and gas very rapidly. Also, to be useable, it must be a condensed material, usually a solid. There are a few liquid explosives such as nitromethane but they are rarely used.

Fuel–Oxidizer Mixtures

These are explosives in which the fuel and oxidizer are two distinctly separate materials. Perhaps the most well-known explosive of this type still used today is gunpowder, sometimes called black powder. Gunpowder consists of a mixture of charcoal (essentially carbon) as fuel, sulphur, which sensitizes the mixture, and potassium nitrate (KNO_3), as oxidizer. The detailed reaction of this mixture is complicated

with references to the reaction containing up to 14 different products. However, in simple terms the carbon is oxidized (burned) by the oxygen held in the nitrate group of the potassium nitrate liberating in excess of 3000 k J kg^{-1} of heat energy and carbon dioxide together with some carbon monoxide. Further gas production comes from the nitrogen in the potassium nitrate forming nitrogen gas (N_2) liberating a total gas output of approximately 300 l/kg^{-1}. A simplistic reaction of black powder is as follows:

$$4KNO_3 + 7C + S \rightarrow 3CO_2 + 3CO + 2N_2 + K_2CO_3 + K_2S$$

This concept has led to the production of many types of explosive. The vast majority of pyrotechnic compositions are mixtures of this type. Examples of pyrotechnic mixtures and their uses are given in **Table 1**. In more recent times fuel–oxidizer mixtures have been used to produce commercial blasting explosives. The first to be produced and still in use today is known as ANFO (ammonium nitrate fuel oil). The hydrocarbon fuel oil (diesel) is oxidized by the oxygen in the nitrate to form carbon dioxide and water. However, unlike black powder where approximately 60% of the products are solids, in this case with the use of ammonium nitrate (NH_4NO_3) essentially all of the products of reaction are gases.

How rapidly these mixtures can react will depend on the intimacy of mixing. This can be controlled by various methods including particle size of the components and mixing method. The speed of burning of pyrotechnic compositions is often controlled by particle size. The ammonium nitrate used to make ANFO is formed into porous 'prills' to give very good absorption of the liquid diesel.

The ability of an explosive to do work will depend on both the amount of heat energy and gas volume produced. Obviously this will vary from one type of explosive mixture to another, however, since this is a chemical reaction between components another major factor affecting the performance of a particular composition will be the ratio of the components. For example, when using essentially pure ammonium nitrate, the optimum mixture for ANFO is 94% ammonium nitrate and 6% diesel.

Explosive Molecules

During the nineteenth century the developing science of chemistry began to create molecular species with explosive properties. These molecules contain atoms that act as fuels, i.e. carbon and hydrogen, but also contain nitro groups (-NO_2) similar to the nitrates described in the mixtures above. Typical examples of these molecular types are shown in **Fig. 1**. There are three basic structural types: nitro compounds containing the C—NO_2 grouping; nitrate esters containing C—O—NO_2 and nitramines containing N—NO_2.

These could be seen as the most intimate 'mixture' of the fuel and oxidizer and will allow for very fast

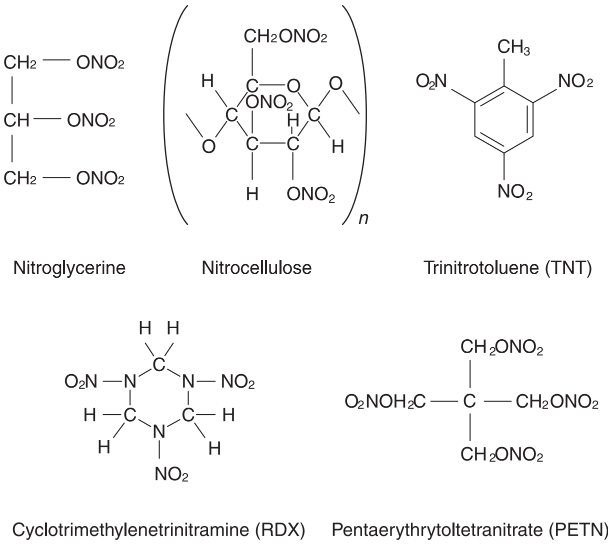


Figure 1 Some typical explosive molecules.

Table 1 Pyrotechnic mixtures

	Fuel	*Oxidizer*	*Other*	*Use*
1	Boron	Barium chromate		Delay
2	Aluminum	Iron oxide		Molten iron for welding
3	Magnesium	Sodium nitrate	Lithographic varnish	Illuminating white light
4	Magnesium	Strontium nitrate	Potassium perchlorate	Signaling red light
5	Magnesium	Barium nitrate	Potassium perchlorate	Signaling green light
6	Magnesium	Teflon (PTFE)	Viton binder	Infrared decoy flare
7	Aluminum	Potassium perchlorate		Thunderflash
8	Aluminum	Zinc oxide	Hexachloroethane	Screening smoke
9	Silicon	Lead oxide		Priming/igniting
10	Magnesium	Potassium nitrate	Resin binder	Igniter

reactions making them superior to explosive mixtures. However, one limitation is that, being molecules, the ratio of fuel to oxidizer is fixed for that particular molecular formula. Thus, nitroglycerine has the formula $C_3H_5N_3O_9$ in which there is sufficient oxygen (nine atoms) to react with the three carbon atoms and five hydrogen atoms to completely oxidize them to carbon dioxide and water:

$$C_3H_5N_3O_9 \rightarrow 3CO_2 + 2.5H_2O + 1.5N_2 + 0.25O_2$$

In fact there is a slight excess of oxygen in this molecule and therefore it is said to have a positive oxygen balance. However, trinitrotoluene has the molecular formula $C_7H_5N_3O_6$ in which there is a significant deficit of oxygen in relation to the fuel content. This is a negative oxygen balance and may be so great, as indeed it is in trinitrotoluene (TNT), that a proportion of the products is carbon giving a black cloud on explosion. The following equation shows the products of a TNT detonation based on an assessment of reaction given by Kistiakowsky and Wilson.

$$C_7H_5N_3O_6 \rightarrow 3.5CO + 3.5C + 2.5H_2O + 1.5N_2$$

The one exception to the constancy of fuel to oxygen ratio in explosive molecules is nitrocellulose. As can be seen in Fig. 1, there are three sites on each of the cellulose subunits where the nitro group can be attached. The reaction between cellulose, where these three sites begin as alcohol groups, and nitric acid tends not to go to completion where all sites have been nitrated. Fully nitrated cellulose as shown would have a nitrogen content of 14.14% by weight, whereas for an average of two of the three sites to be nitrated this figure would be 11.11%. The practical upper limit for nitration is 13.4% nitrogen by weight.

To form gaseous products does not necessarily require the formation of carbon dioxide as there is another gaseous oxide of carbon when it is partially oxidized, i.e. carbon monoxide. If the reaction of cyclotrimethylenetrinitramine (RDX) is considered with a formula of $C_3H_6N_6O_6$ then there is sufficient oxygen to convert all of the fuel elements into gas, each molecule creating three molecules of carbon monoxide (CO) and three molecules of water (H_2O). In reality, explosive formulations based on these types of molecule which are used mainly for military purposes have a negative oxygen balance and therefore will have significant quantities of carbon monoxide in the gaseous products. This has been the cause of accidental deaths when explosives have been used in confined spaces and the gases can accumulate. Commercial explosive being mixtures will attempt to eliminate as far as possible the formation of carbon monoxide by having formulations with an oxygen balance near to zero. This is particularly important for tunneling or other underground work.

The performance of an explosive in its ability to do work is determined by the amount of heat and gas produced in the initial reaction. Values are given in **Table 2** for a range of explosive molecules. It should be noted that, when mixing two explosive types together to make a formulation, the values for heat and gas production cannot be found simply by proportion. The products of reaction depend on the ratios of carbon and hydrogen to oxygen within the actual formulation not just within each component.

Subsequent reactions of the initial products can and do occur. For example, gun propellants have a negative oxygen balance and as there is only a very small amount of oxygen in the breech then the hot gaseous products will remain without further reaction until they exit the muzzle of the gun. Here they mix rapidly with a large volume of air and since the gases contain carbon monoxide and often hydrogen, both highly flammable, a secondary burning reaction occurs. This is the main cause of muzzle flash which is seen best for large caliber guns.

Table 2 Heat and gas production from explosive molecules[a]

Explosive	*Heat of explosion* (*Q*) (kJ kg^{-1})	*Gas volume* (*V*) (l kg^{-1})	*Comments*
Ethyleneglycoldinitrate	6730	740	Zero oxygen balance
Nitroglycerine	6275	740	
Nitrocellulose	4132	875	13.3% N
Trinitrotoluene	4080	790	TNT
Cyclotrimethylenetrinitramine	5130	908	RDX or Hexogen
Cyclotetramethylenetetranitramine	5130	908	HMX or Octogen
Pentaerythrytoltetranitrate	5940	790	PETN
Nitroguanidine	2680	1077	Picrite
Picric acid	3745	790	Obsolete, standard

[a] Note: These are representative values as different highly reputable sources give variable values, for example, values for RDX heat of explosion have been found to vary from 5100 to 5940 kJ kg^{-1}

The Burning Explosion Process

Certain explosive types such as propellants create energy and gas release through a surface reaction. At the surface the chemistry of combustion occurs where the fuel present in the material is being oxidized by the source of oxygen within the material. Thus, it is a self-contained energy system where a set of complex chemical reactions ultimately produces the combustion products and heat, which leaves the burning surface. However, since this burning surface is a heat source, some of the heat will conduct into the bulk of the explosive and raise the layer immediately below the burning surface to a higher temperature. This will cause ignition of this next layer at which time the layer creating the heat for ignition will have turned to gaseous products. This is shown pictorially in **Fig. 2**.

Thus, the flame front will move in to the burning explosive at right angles to the surface so a sphere as shown will simply decrease in radius. The speed that the flame front moves is known as the linear burning rate, however, this alone cannot predict the mass burning rate, i.e. the rate at which a given mass of explosive is turned to heat and gas. Being a surface phenomenon, the surface area, which is burning, must affect the mass burning rate, thus at any given moment:

$$\frac{\mathrm{d}m}{\mathrm{d}t} \propto A \times r$$

where A is surface area and r is the linear burning rate. This means that if the burning particle changes shape and therefore changes surface area, the mass burning rate will change during the explosion process.

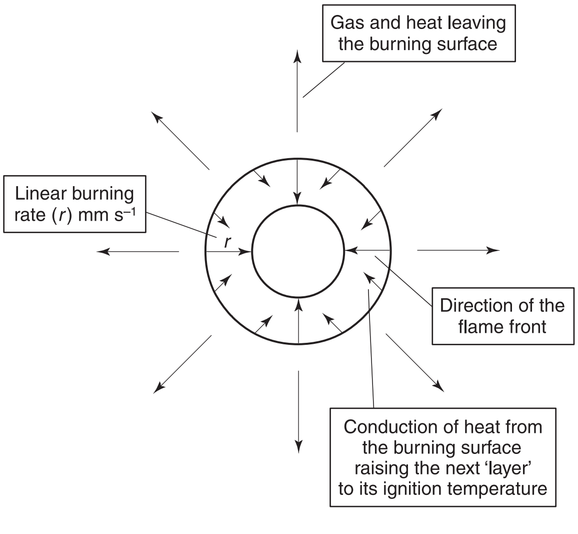


Figure 2 The burning process for a sphere of explosive.

In real propellant systems this is taken into account in two ways. Firstly, the size of the particles known as 'grains' can control the initial surface area for a given mass of explosive. The smaller the grains, the larger the surface area and so small caliber weapons such as small arms will have small masses of propellant within the cartridge case and will have small grains. However, a large caliber gun such as a 155 mm howitzer using propelling charges with masses measured in kilograms uses much larger grain sizes; otherwise the initial mass burning rate based on initial surface area would be so great that the pressure would rise in the breech to dangerous levels.

The second method of control is to choose grain geometry such that the change in surface area can influence the mass burning rate. For example, if the grain is solid such as a sphere or flake then the surface area will decrease during the burning process. However, if the grain is a cylinder with axial multiperforations, then since all available surfaces are burning the internal holes will increase in surface area. When there are sufficient holes to overcome the decrease in surface area of the outer surface of the cylinder the overall effect is for the surface area to increase during the burning process. A grain shape giving an increase in surface area is known as 'progressive', for a decrease 'degressive' and if the surface area remains constant 'neutral'.

There is a further factor that influences the mass burning rate; this is the main reason that burning explosives can be used to do work. In the last decade of the nineteenth century a French scientist called Vieille discovered that the linear burning rate of an explosive was affected by pressure above the burning surface. As the pressure increases, the linear burning rate increases, which is what occurs when propellants are used in the breech of a gun or in a rocket motor. Thus, confinement of a burning propellant will increase the mass burning rate significantly. Typically, the linear burning rate of a gun propellant in the open at atmospheric pressure will be about $5\,\mathrm{mm\,s^{-1}}$, whereas in the breech of a gun it can rise to $400\,\mathrm{mm\,s^{-1}}$ at peak pressure. The relationship is often given as:

$$r = Bp^a$$

where a is the pressure index and p is pressure. Each explosive material will have unique values of a and B.

This concept is used in noise-producing pyrotechnics such as signaling maroons, display fireworks or military training devices. This is also the mechanism by which the pipe bomb works by confining a burning explosive such as black powder, improvised mixture or gun propellant in a metal tube and igniting.

The Detonation Process

Detonation is a completely different mechanism to the burning process above. The burning process proceeds by heat conduction, whereas a detonation proceeds by the passage of a shock wave through the explosive charge. Being a shock wave it must be traveling at a velocity equal at least to the speed of sound within that material. As the speed of sound in explosives is approximately $1800\,m\,s^{-1}$, this is the minimum velocity at which a detonation occurs.

The process and energy outputs are shown diagrammatically in **Fig. 3**. Once an explosive charge has been initiated to detonation, a shock wave passes through the explosive causing the explosive chemistry to begin. The main mechanisms are the adiabatic compression of microscopic voids and crystal boundary effects creating hot spot generation as the intense pressure of the shock wave passes. The energy release and gas production occur in a reaction zone immediately behind the detonation shock front. The reaction zone maintains pressure on this shock front creating a steady-state speed known as the velocity of detonation, values of which are given for a range of explosives in **Table 3**. The thickness of the reaction zone depends among other things on the type of explosive related to how fast the chemistry can occur and on the size of the charge but normally it will be a few millimeters.

The shape of the shock wave front depends on charge boundaries. In a theoretical charge of infinite size and a point initiation, the shock front would spread out radially. However, most detonating charges, whether they are the filling of a warhead or a blasting charge in a borehole, essentially have a narrow cross-section relative to their length. Edge effects tend to decrease the curvature from that expected in a theoretical radial spread. Thus, in real charges the shock front will be slightly curved.

The intense pressure of the detonation shock wave will act on anything touching, or in very close proximity to (~1 charge diameter), the explosive charge. Values for these pressures are given in Table 3 and from these values (200 kbar = 20 GPa = 2.9 E+6 psi) it is not surprising that the shock wave will cause a shattering effect. This is called the brisance of a detonating explosive and from cased charges as in military warheads, shells, grenades, etc. creates the fragmentation. In blasting it creates an area known as the crush zone immediately surrounding the borehole. The factors that affect the detonation pressure are the velocity of detonation (D $m\,s^{-1}$) and charge density (ρ $g\,cm^{-3}$). Using these units then the detonation pressure (p kbar) can be calculated from:

$$p = \rho D^2 \times 2.5\ \mathrm{E} - 06$$

An explosive charge will detonate without confinement. Indeed, confinement has little effect on the process for large charges and it is only when charge diameters become small that confinement begins to have an effect. For a cylindrical bare charge, at large diameters a decrease in diameter will not change the velocity of detonation. As the diameter begins to be significantly reduced a point is reached where below this diameter the velocity of detonation decreases with decreasing charge diameter. Finally a diameter

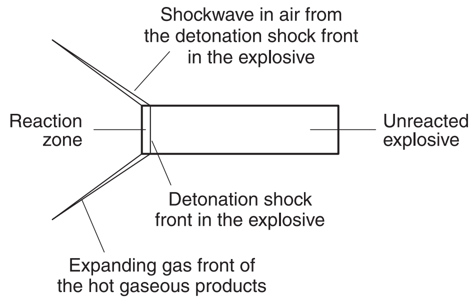


Figure 3 The detonation process for a cartridge of explosive.

Table 3 Detonation parameters of explosive molecules

Explosive	*Typical density* (ρ) (*g cm*$^{-3}$)	*Velocity of detonation* (*D*) (*m s*$^{-1}$)	*Detonation pressure* (*p*) (*kbar (GPa)*)	*Power index*[a]
EGDN	1.48	7300	197 (19.7)	168
Nitroglycerine	1.59	7600	230 (23.0)	157
Trinitrotoluene	1.65	6900	196 (19.6)	122
RDX	1.82	8750	348 (34.8)	157
HMX	1.9	9100	393 (39.3)	157
PETN	1.76	8400	310 (31.0)	159
Nitroguanidine	1.71	8200	287 (28.7)	98
Picric acid	1.767	7350	239 (23.9)	100

[a] Calculated from the data in Table 2.

is reached for a bare charge below which the charge cannot support a detonation. This is called the critical diameter. Confining the charge reduces the diameters at which these phenomena occur since they are caused by energy loss from the sides, which affects the ability of the reaction zone to keep the pressure on the detonation shock wave.

If the detonation shock wave passes directly into the surrounding air it loses energy very rapidly by compressive heating effects and at a position approximately two charge diameters behind the reaction zone is caught by the expanding gaseous products. These expanding gases in the early stages of expansion have a velocity measured in 1000s m s^{-1}. This rapid expansion pushes the surrounding air so violently that another pressure wave is formed and becomes a shock wave in air. This is the blast wave from a detonation which travels away from the explosion and reduces in peak pressure and impulse with increasing distance.

The expanding gases create the power of a detonating explosive and as such affect the ability to provide lift and heave for cratering, trenching and quarrying. Quantification of power of an explosive is usually done by comparison with a standard explosive. Military explosives are often compared to a TNT equivalence, thus, if an explosive had a TNT equivalence of 1.2 it would mean that every 1 kg of this explosive would have the lift and heave or blast wave effect of 1.2 kg of TNT. Another method is to calculate the equivalence of heat and gas production with that from a standard explosive. If the standard is given a scaling figure then all explosives can be compared with each other on this index. One such standard is picric acid and power index values using this standard are given in Table 3.

Initiation of the Explosion Processes

Virtually all explosives, except pyrotechnics, will both burn and detonate, the type of event depending on the method of initiation. There are a small number of pyrotechnics that will detonate. As already mentioned some explosives such as propellants are designed to burn and others to detonate and thus the initiation method must be specific and reliable.

To maintain safety it is not surprising that a bulk explosive charge of any type should be as insensitive as possible. Equally, the applied stimulus to initiate an explosive cannot be excessive. Thus, practical initiation for both the burning and detonation processes is achieved through the use of an explosive train.

The first element of an explosive train is a primary explosive that is very sensitive and present in small quantity, usually 0.1–0.5 g. Since the main charge may be quite insensitive the output of this primary explosive is usually amplified using a booster, which in turn will initiate the main charge. For small systems such as the propelling charge of a small arms bullet the booster may not be present. The primary explosive used for igniferous (burning) initiation is usually lead styphnate, often mixed with a pyrotechnic material to enhance the flash of flame produced. In practical systems the normal stimuli used to initiate the primary explosive are percussion as in a bullet percussion cap or electrical heating of a wire that is coated with the primary explosive.

The primary explosive used in detonating systems is lead azide and is particularly useful in that it undergoes a spontaneous transition from burning to detonation under all conditions. This allows simple methods to be used to initiate the lead azide pellet in the explosive train to burning, a flash of flame from a fuze or from an electrically initiated match-head, and then within a fraction of a millimeter convert in to a shock wave and thus be detonating. If the lead azide is placed next to a booster pellet this will undergo a shock-to-detonation process where the shock output of the lead azide detonates the booster and in turn the booster pellet will do the same to the main charge.

The booster in a detonation train is likely to be PETN or RDX based and is more sensitive than the main charge. However, both the booster and main charge are classed as secondary explosives to differentiate them from the very sensitive primary explosive. In military warheads the primary explosive pellet is kept physically separated from the booster until the fuze is armed. Fuze design will not normally allow this to occur until the warhead has left the launch platform; environmental effects are often used to trigger the arming mechanism. For example, many air-dropped weapons use airflow to spin a vane, which must spin for a certain number of revolutions before the primary explosive can move into position. An artillery shell must experience both the set-back force of acceleration and the spin imparted in the barrel before the fuse will arm.

In the commercial sector for blasting etc. the initiation method is the demolition detonator. This is a thin metal tube 6–8 mm in diameter, usually constructed of aluminum or copper in which there is a base charge as booster, often PETN and a lead azide pellet. In the electrically initiated version the primary explosive is initiated by an electric fuze head. This is inserted into a bulk charge such as a cartridge of blasting explosive. A diagrammatic representation is given in **Fig. 4**.

See also: **Explosives**: Analysis; Commercial; Military; Bomb-scene Management.

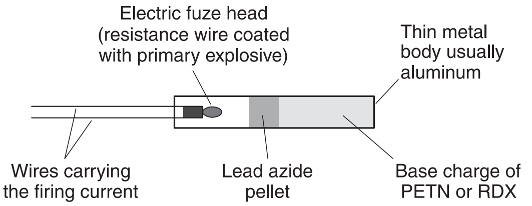


Figure 4 Structure of an electric demolition detonator.

Further Reading

Akhavan J (1998) *The Chemistry of Explosives*. Cambridge: The Royal Society of Chemistry.

Bailey A and Murray SG (1989) *Explosives, Propellants and Pyrotechnics*. London: Brassey's (UK) Ltd.

Bulusu SN (ed) (1990) *Chemistry and Physics of Energetic Materials*. Dordrecht: Kluwer.

Brown GI (1998) *The Big Bang: a History of Explosives*. Stroud, Sutton.

Conkling JA (1985) *Chemistry of Pyrotechnics: Basic Principles and Theory*. New York: Marcel Dekker.

Cooper PW (1996) *Explosives Engineering*. New York: Wiley-VCH.

Kohler J and Meyer R (1993) *Explosives*, 4th revised and extended edn. Weinheim: VCH.

Olah GA and Squire DR (eds) (1991) *Chemistry of Energetic Materials*. San Diego: Academic Press.

McLain JH (1980) *Pyrotechnics: From the Viewpoint of Solid State Chemistry*. Philadelphia: Franklin Institute Press.

Suceska M (1995) *Test Methods for Explosives*. New York: Springer.

Urbanski T (1964) *Chemistry and Technology of Explosives*. Oxford: Pergamon Press.

Military

S G Murray, Cranfield University, Royal Military College of Science, Shrivenham, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0509

Introduction

To secure the best value from this article it is recommended that the article entitled 'Mechanism of Explosion' be read first.

Until the nineteenth century the military use of explosives was limited to black powder, also known as gunpowder when used in muskets and cannon, as the only available material. The advent of nitrated molecules opened the way to explosives with far better energetic properties but also the ability to create detonations. Thus, two families of military explosives were born: those that burned explosively and were used to replace the obsolete gunpowder and those that could detonate, which became known as 'high explosives'. The burning type of explosives were called 'low explosives' but in general were for use in guns and later rockets and so became classified as propellants.

Within the various sections below, consideration is given to some historic military explosives as old ammunition is still found in collections or recovered from battlefields such as those in France and Belgium from the 1914–1918 conflict.

High Explosives

Historical

The discovery of the mercury fulminate detonator in the 1860s allowed the full potential of the newly developed nitroglycerine/nitrocellulose explosives to be exploited. However, these new formulations invented by the Nobel family were not suitable for military use in explosive ordnance. However, at the turn of the nineteenth century other researchers were applying the concept of detonation to some different explosive molecules, some of which had been known for almost 100 years. One of the first to be developed into fillings for ordnance was picric acid (2,4,6-trinitrophenol), either as a pure material or mixed with dinitrophenol (mixed isomers) to lower the melting point of the mixture to aid melt casting.

At the same time the explosive 2,4,6-trinitrotoluene (TNT) (**Fig. 1A**) was also being developed and was found to be superior to explosives based on picric acid. The disadvantages of picric acid were that it reacted with metals to give very impact-sensitive compounds and on the other hand was at times difficult to detonate reliably. Thus, although picric acid formulations survived until the 1940s, after this time they became completely obsolete. The use of TNT was highly successful, not only as a pure filling but, by the end of World War I, as a mixture with ammonium nitrate to give the explosive known as amatol which could contain as much as 80% ammonium nitrate. The main driving force towards the use of amatol was that the requirement for explosives during the war could not be met solely by production of TNT, whereas ammonium nitrate could be produced in large quantities thereby supplementing the TNT when mixed in a formulation.

By the beginning of World War II, research into

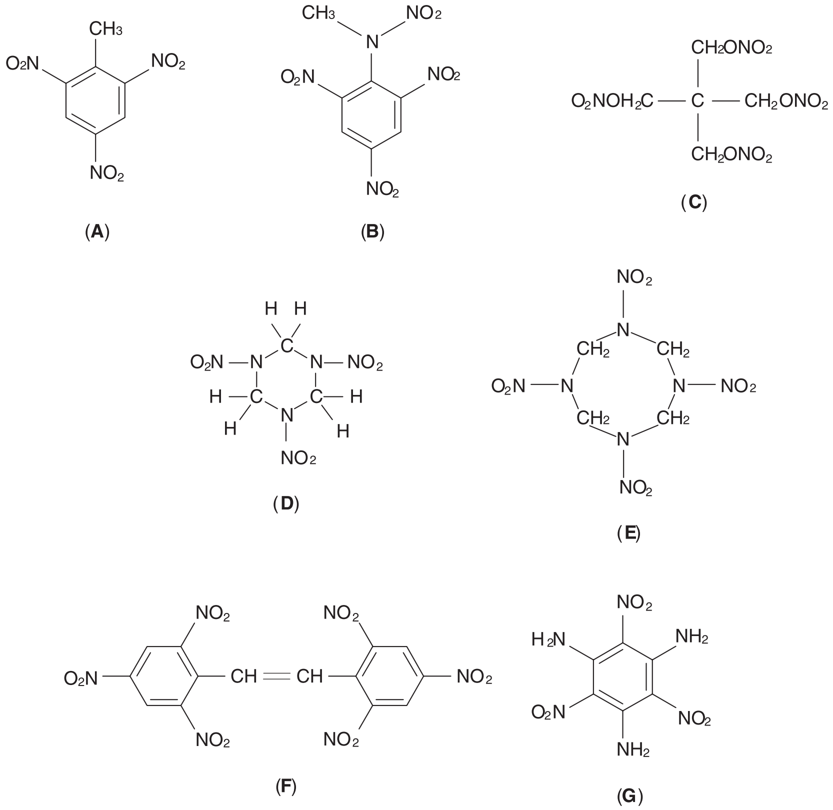


Figure 1 Structure of 2,4,6-trinitrotoluene (TNT) (**A**), tetryl (**B**), pentaerythrytoltetratrinitrate (PETN) (**C**), cyclotrimethylenetrinitramine (RDX) (**D**), cyclotetramethylenetetranitramine (HMX) (**E**), HNS (**F**), TATB (**G**).

other explosives had identified another group of explosive molecules that could be used for the filling of ordnance. One of these was tetryl (Fig. 1B), which had ideal explosive properties to be a booster in a detonating explosive train. Thus, from that time tetryl has been used worldwide as the most commonly used booster explosive. It is only recently that the toxicity of this explosive, causing severe dermatitis, has provided the impetus to find replacements. Even so, it is still in use today. The other major developments were based around three explosive molecules, pentaerythrytoltetranitrate (PETN) (Fig. 1C), cyclotrimethylenetrinitramine (RDX) (Fig. 1D) and cyclotetramethylenetetranitramine (HMX) (Fig. 1E). These, together with TNT are the basis for all modern explosives.

Requirements for a military high explosive

Military high explosives are required to provide the following effects:

- Fragmentation of metal casings as in grenades, shells, mines, etc.;
- Blast (pressure wave in air created by the expanding gases from the explosion);
- Underwater bubble pulse for mines and torpedoes;
- Lift and heave for cratering;
- Shaped charge phenomenon.

Fragmentation is created by the shock wave in the detonating explosive breaking the casing, a phenomenon called brisance, and then accelerating the fragments. Thus, to achieve this requires an explosive with a reasonably good detonation shock pressure. Blast, underwater bubble pulse and lift and heave are all created by the expanding gases from the detonation and are not affected by the detonation shock pressure. Thus, it would be beneficial to increase the gas expansion effects at the expense of brisance. The shaped-charge phenomenon is the creation of a metal jet from a hollow metal cone backed by the explosive (**Fig. 2**). The jet tip can be traveling at a velocity of approximately $10\,000\,m\,s^{-1}$ and can penetrate armour, the design requirement of this type of ordnance. The best explosive for this is one with the highest detonation pressure. The gas expansion is not relevant to the jet formation.

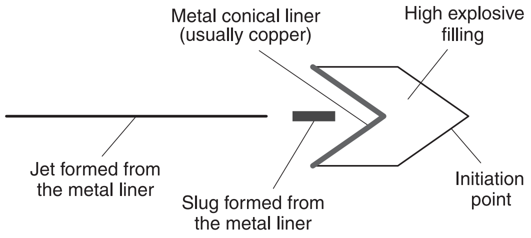


Figure 2 Shaped-charge warhead design and function.

The above requirements are based on the explosive performance. There are other requirements that are important for military explosives, which must also be considered. These are:

- Safety during storage, transportation and use;
- Long shelf life;
- Reliability;
- Ease of filling into ordnance.

This has led to a series of compositions capable of fulfilling all of these requirements based almost entirely on the explosive molecules RDX, HMX and TNT together with a limited use of PETN and a few highly specialized compounds such as HNS (Fig. 1F) and TATB (Fig. 1G). The performance of these as pure compounds together with some safety related data are given in **Table 1**.

Formulations

Table 1 shows that RDX, HMX and PETN are all capable of providing very high shock pressures with HMX being the best. However, PETN is significantly more sensitive than RDX and this has led to the preference of RDX over PETN where safety is an issue in such uses as artillery and cannon rounds where the set-back forces on launch may cause premature initiation. As the performances of RDX and PETN are very similar, many countries have chosen to use RDX in preference to PETN in most other cases. HMX is slightly more sensitive than RDX but significantly more expensive. It is used where very high shock pressures are required.

Thus, military high explosives are based on RDX, HMX and to a lesser extent PETN to provide the necessary performance. However, for safety it is considered that they are too sensitive to use on their own except when used as booster pellets. To provide a formulation that is acceptable for both safety and performance they are mixed with either TNT or an inert phlegmatizer, both of which reduce the sensitivity. The choice of these binders is governed to some extent by their ability to provide practical methods of filling ordnance. For example, the presence of TNT aids the filling process due to its low melting temperature (see Table 1). The TNT is melted with low pressure steam and held at ~90°C. The other components are then added and mixed to form a slurry, which then can be poured into the ordnance. The only drawback to this process is that the TNT contracts by 10% as it solidifies and this necessitates the use of equipment that can give a filling with no voids. This is important as the presence of voids in the filling of an artillery shell may cause premature initiation of that filling as the round is fired. This is caused by rapid compression of the air-filled voids creating adiabatic heating, thus igniting the explosive and causing an explosion in the barrel of the gun.

The other method used to desensitize the RDX, HMX or PETN is with an inert material. This may be a wax that has a melting point around 140°C and can be used to cast/press into a warhead. Another method

Table 1 Performance data for some compounds used in military explosive formulations.

Compound name	*Melting point (°C)*	*Detonation velocity (m/s^{-1})*	*Detonation pressure (kbar)*	*Power[a]*	*Impact sensitivity (N m)*
RDX	205[b]	8700 @ 1.77 g cm^{-3}	338 @ 1.77 g cm^{-3}	480	7.5
HMX	285	9110 @ 1.89 g cm^{-3}	390 @ 1.89 g cm^{-3}	480	7.4
PETN	140	8260 @ 1.76 g cm^{-3}	335 @ 1.77 g cm^{-3}	523	3
TNT	80.9	6900 @ 1.60 g cm^{-3}	210 @ 1.63 g cm^{-3}	300	15
TATB	452[c]	7760 @ 1.88 g cm^{-3}	291 @ 1.88 g cm^{-3}	250[d]	50

[a] Expansion volume for a 10 g sample in the lead block test.
[b] With decomposition.
[c] Decomposes above 325°C if heated slowly.
[d] Calculated value as a 10 g sample does not fully detonate.

is to mix with a plasticizer-type material, which imparts a moldable consistency. A third method is to mix with a plastic giving a range of plastic bonded explosives (PBX). And lastly, a recent development has been to set the explosive powder in a cross-linked polymer. These types are described below.

Moldable explosives

Moldable explosives are sometimes known generically as plastic explosives. This may lead to some confusion, as there are PBX and cross-linked polymer-bound explosives. However, the term 'plastic explosive' is generally taken to mean one that is capable of being molded into shape by hand and used for demolition purposes. Some of the more well-known examples of this type of formulation are C4 (USA), PE4 (UK) and the Semtex family A, H and 10 (Czech Republic). The compositions are given in **Table 2**.

Castable explosives

As described above, molten TNT can be utilized as a vehicle for casting formulations that include solid powders. These powders include the explosives RDX, HMX, PETN and tetryl together with a variety of other materials such as ammonium nitrate, barium nitrate and aluminum. This has led to a range of explosives developed for various reasons.

The first castable explosives based on TNT were the amatols in which the TNT was mixed with ammonium nitrate. The main driving force behind this was the shortage of TNT during wartime periods. Since ammonium nitrate is oxygen rich and TNT oxygen deficient, the formulations gave a more oxygen balanced mixture with good explosive properties. Similarly, a range of explosives was developed based on barium nitrate mixed with TNT, the baratols. A range of formulations based on tetryl/TNT mixtures called tetrytols also developed. All of these are now obsolete and were replaced by mixtures containing RDX, HMX and PETN. The RDX/TNT mixtures that developed ranged from 50/50 to 75/25 RDX/TNT and are known by the US name of cyclotols. The most commonly used type is nominally 60/40 RDX/TNT, often called Composition B. Those containing HMX are known as octols with high percentages of HMX and those with PETN the pentolites.

One other major component mixed into these castable explosives is aluminum powder. This undergoes further chemical reaction with the gaseous products of explosion (CO_2, H_2O and CO) to form aluminum oxide by extraction of oxygen from the gases at the temperature of the explosion. This creates more heat and at modest aluminum contents no loss of gas volume. The overall effect is to increase the power of the explosive, that is the gas expansion effect. In the UK these RDX/TNT/aluminum formulations are the Torpex series of explosives and in the US, H-6 and HBX-1 to -3. Tritonal contains only TNT and aluminum. An obsolete explosive of this type is Ammonal containing ammonium nitrate/TNT/aluminum.

Compositions for the castable explosives are given in **Table 3**.

Table 2 Components of some typical moldable explosives.

Name	*RDX (%)*	*PETN (%)*	*Desensitizer*
C4	91		Di(2-ethylhexyl) sebacate Polyisobutene Motor oil
PE4	88		Lithium stearate Paraffin oil
Semtex A		83.5	Rubber Oil
Semtex H	85.5 (mixture)		Rubber Oil
Semtex 10		85	Rubber Dibutylformamide

Table 3 Formulations of some typical castable military explosives.

Name	*TNT (%)*	*RDX (%)*	*Other components (%)*
Amatol	20		Ammonium nitrate 80
Baratol	24		Barium nitrate 76
Composition B	39.5	59.5	Wax 1
Cyclotol	25	75	
H-6	30	45	Aluminum 20 Wax 4.5 Calcium chloride 0.5%
Minol	40		Ammonium nitrate 40 Aluminum 20
Octol	25		HMX 75
Pentolite	50		PETN 50
Tritonal	80		Aluminum 20

Plastic bonded explosives (PBX)

Plastic bonded explosives utilize plastics to bind and desensitize RDX and HMX or to act simply as a binder with TATB, an explosive compound that is very insensitive. A typical preparation of a PBX would be to suspend the explosive powder, e.g. RDX, in water and add the plastic dissolved in an organic solvent that does not dissolve the explosive but wets the surface. The plastic is then precipitated from the solution either by addition of another solvent or by evaporation of the solvent that dissolves the plastic. Thus, the plastic coats the explosive crystals and forms what is called the molding powder. This powder can then be pressed directly into the ordnance or isostatically in a rubber membrane followed by machining to shape. This pressing is carried out at high vacuum and achieves up to 97% of the theoretical maximum density. Details of the plastics used are given in **Table 4**.

There are several reasons why this technique is favored over TNT casting to provide formulations. One is to obviate the necessity to have TNT in the filling as the low melting point may be disadvantageous and also TNT is prone to cracking. However, the main reason is that it is possible to obtain formulations with very high RDX or more significantly HMX, content. To produce warheads that require very high detonation pressure it is necessary to have such high explosive loading but still be acceptable from a sensitivity standpoint. This is achievable with PBX. The majority of PBX formulations are from the US Armed Forces with designations LX-, PBX-, PBX(AF), PBXC-, PBXN- and PBXW- followed by a number. Some examples are given in **Table 5**.

A recent development in the manufacture of PBX has been the use of thermosetting polymer systems to form a matrix of solid powders in the polymer by polymerization after a mixture of ingredients plus liquid prepolymer has been formed. The polymerization is induced by gentle heating and the polymer forms without any volume change and can thus occur with the prepolymer mixture already in the warhead. One such system utilizes the reaction between hydroxy-terminated polybutadiene crosslinked with a di-isocyanate producing a polyurethane. The products are rigid, have the physical properties controlled by choice of polymer system and cannot be melted by heat.

Other chemicals that might be present in PBX formulations are energetic plasticizers. These include mixtures of di- and tri-nitroethylbenzenes, bisdinitropropylacetal/bisdinitropropylformal (BDNPA/F) and bis(2-fluoro-2,2-dinitroethyl)formal (FEFO). At times it appears that an explosive is classed as a PBX by the presence of these energetic plasticizers. Examples can be found within the PBXN formulations of the US Department of Defense. Also, there are a few aluminized PBX formulations.

Table 4 Some polymers used in the manufacture of PBX.

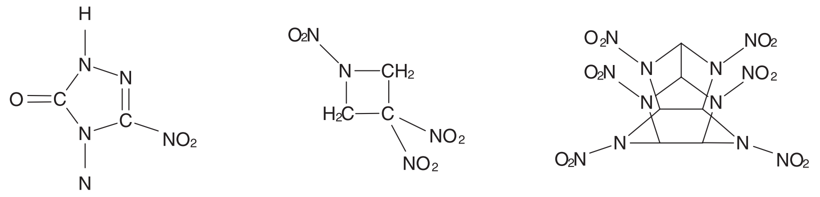
Name	*Description*
Viton A	Vinylidine fluoride/hexafluoropropylene copolymer 60/40 wt%
PolyDNPA	2,2-Dinitropropyl acrylate polymer
Estane	Polyurethane
Kel-F 800	Chlorotrifluoroethylene/vinylidine fluoride copolymer (3:1)
Polystyrene	Polymerized styrene monomer

Table 5 Some typical PBX formulations.

Name	*Explosive component*	*Plastic (%)*
LX-09-0	HMX 93	PolyDNPA 4.6 FEFO 2.4
LX-10-0	HMX 95	Viton A 5
LX-11-0	HMX 80	Viton A 20
LX-14-0	HMX 95.5	Estane 4.5
LX-15	HNS 95	Kel-F 800 5
LX-17-0	TATB 92.5	Kel-F 800 7.5
PBX-9205	RDX 92	Polystyrene 6 DIOP 2
PBX-9604	RDX 96	Kel-F 800 4

Recent developments

Irrespective of the method of formulation, there are some recent additions to explosive compounds that will be used in future explosives. Probably the two most important additions are 3-nitro-1,2,4-triazol-5-one (NTO) and hexanitrohexaazaisowurtzitane (HNIW or CL20). These are currently being developed into ordnance fillings. Other new contenders include 1,3,3-trinitroazetidine (TNAZ). The chemical structures are shown in **Fig. 3**.



NTO TNAZ CL20

Figure 3 Some new compounds of relevance to military high explosives. NTO, 3-nitro-1,2,4-triazol-5-one; CL20, hexanitrohexa-azaisowurtzitane; TNAZ, 1,3,3-trinitroazetidine.

Propellants

Gun propellants

The discovery of nitroglycerine and nitrocellulose led to the invention of 'smokeless powders' by workers such as Vieille, Nobel, Abel and Dewar. These were far superior to the gunpowder that they replaced. Thus, at the beginning of the twentieth century the use of gunpowder for military use had ceased.

The early forms of these new propellants were as extruded strands sometimes called cords, indeed, they were known generically as cordites. Otherwise they were spheres or flakes, the propellant powders, an example being ballistite. As the science of gun ballistics rapidly developed and these new propellants were used in large caliber guns, the geometries of the propellant pieces, or 'grains', developed leading to tubular, stick and multitubular granular propellants. Some of these shapes are shown in **Fig. 4**.

Propellant formulations have developed since the early days initially into three basic types and more recently with the addition of a fourth type. They are differentiated by the explosive components:

- Single base propellant which contains nitrocellulose (NC);
- Double base propellant which contains NC and nitroglycerine (NG);
- Triple base propellant which contains NC, NG and nitroguanidine (picrite);
- High energy propellant which contains NC, NG, picrite and RDX (or HMX).

Nitrocellulose can be nitrated to differing amounts and thus can have different energies but, being polymeric, one of its main functions is to provide physical properties that resist grain cracking and also gels with NG if present. The NG is a highly energetic explosive and thus its presence increases the energy of the propellant formulation. Nitroguanidine (**Fig. 5**) contains a large percentage of nitrogen in the molecule. It is present in triple base propellant in significant quantities, typically 55%. Thus, the gases that exit the barrel of the gun are extremely rich in nitrogen gas. This prevents the other gaseous products rich in hydrogen and carbon monoxide from spontaneous ignition thus acting as a muzzle flash suppressant. The RDX in high energy propellant increases the performance without an unacceptable increase in flame temperature. High flame temperatures lead to barrel erosion.

These formulations also contain additives to aid safety, performance and manufacture. The most important of these is the stabilizer. Both NC and NG are inherently unstable and very slowly decompose to release acidic species based on nitrogen and oxygen. If left unchecked, these decomposition products attack unreacted molecules thus accelerating further decomposition. To prevent such a runaway situation additives such as diphenylamine or ethyl centralite

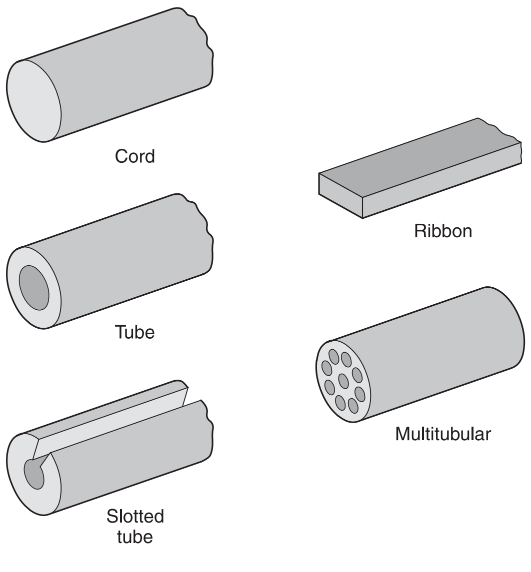


Figure 4 Grain shapes of gun propellant.

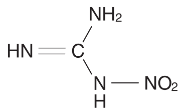


Figure 5 Structure of nitroguanidine (picrite).

(carbamite) are included in the formulation. They do not prevent the slow decomposition; rather they chemically remove the decomposition products preventing attack on unreacted molecules. Thus, during the life of a propellant this stabilizer additive is slowly consumed. Surveillance programs for stored ammunition monitor the depletion and it is usual to destroy the stock when the stabilizer falls to 50% of the original content.

Other additives include plasticizers such as dibutylphthalate, diethylphthalate and glycerol triacetate (triacetin) to modify physical properties, dinitrotoluene as a moisture repellent and cryolite as another type of flash inhibitor. The small grains as found in small arms propellants are likely to be coated with a graphite glaze. This is to aid filling by acting as a lubricant, to increase moisture resistance, to dissipate electrostatic charge and to modify the initial burning reaction. For the triple base and high energy propellants a dye may be added for color coding.

As for the use of these propellants, they tend to be grouped as follows:

- Single base – small arms and large caliber guns;
- Double base – small arms, machine pistols, small cannon and mortars;
- Triple base – large caliber artillery and tank guns;
- High energy – high velocity tank guns.

Table 6 gives the components of typical examples of these propellants.

Rocket propellants

With the development of the new gun propellants based on NC and NG it was a logical progression that rocket propellants would follow a similar path. Indeed, a range of rocket propellant formulations exist based on these explosives and are known as double base rocket propellants. There are two basic types: extruded and cast.

Extruded rocket motor grains are made from a colloid of nitrocellulose and nitroglycerine. As with the gun propellants, a stabilizer is required, usually carbamite and a plasticizer such as diethylphthalate. A typical formulation might contain over 40% nitroglycerine, a level no longer found in gun propellants due to erosion problems and over 50% nitrocellulose. Other ingredients may be potassium sulphate, carbon powder and wax.

Cast double base rocket motor grains are manufactured by first filling a mold with a casting powder based on nitrocellulose grains containing some of the additives such as stabilizer, plasticizer and a platonizing agent, normally a lead salt. A casting liquid is then added consisting of nitroglycerine mixed with a desensitizer such as triacetin. This is absorbed in to the casting powder creating a solid mass, which is cured for a few days at ~60°C. The platonizing agent influences the relationship between burning rate and pressure such that at the working pressure of the motor, the burning rate is virtually independent of pressure.

The alternative type of rocket propellant is called a composite propellant. This is based on a fuel–oxidizer mixture. The oxidizer is ammonium perchlorate and the fuel a plastic or polymer also acting as binder. Early versions in the 1950s and 1960s were known as plastic composites and contained polyisobutene or polyvinylchloride. A simple process using a plasticizer and heat allowed mixing of the plastic with the oxidizer followed by extrusion. This type of composite propellant is now obsolete.

The preferred composite is often called a rubbery

Table 6 Typical formulations of some gun propellants.

Propellant type	*NC (%)*	*NG (%)*	*Picrite (%)*	*Nitramine (%)*	*Other[a] (%)*
Single base	85–95				DPA 1
					DNT 0–10
Double base	60–90	8–50			EC 1–3
Triple base	16–22	16–22	~55		EC 2–7
					Cryolite 0.3
					K_2SO_4 1–2
High energy[b]	20	20	30	25	EC 2

[a] DPA (diphenylamine), DNT (dinitrotoluene), EC (ethylecentralite).
[b] These are classified and therefore the values given are an estimate by the author.

composite and is based on the concept of polymerizing a liquid prepolymer mixed with the ammonium perchlorate. Several polymer systems have been used including:

- Polyurethane (PU) based on the di-isocyanate/glycol reaction;
- Polybutadiene (PB) polymers using hydroxy- or carboxy-terminated PB with a di-isocyanate;
- Polybutadiene–acrylic acid–acetonitrile (PBAN) with epoxide or aziridine crosslinking.

The favored system appears to be that based on hydroxy-terminated polybutadiene (HTPB).

Various hybrids also exist designed to improve either the performance or the physical properties over the two types described above. Thus, there are formulations containing nitroglycerine, nitrocellulose, ammonium perchlorate, aluminum and even HMX. These are prodigious energy producers with a heat of burning as high as $7700\,J\,g^{-1}$ compared to pure nitroglycerine at $6300\,J\,g^{-1}$.

See also: **Explosives**: Analysis; Mechanism of Explosion; Commercial. **Firearms**: Weapons, Ammunitions and Penetration.

Further Reading

Bailey A and Murray SG (1989) *Explosives, Propellants and Pyrotechnics*. London: Brassey's (UK) Ltd.

Cooper PW (1996) *Explosives Engineering*. New York: Wiley-VCH.

Kohler J and Meyer R (1993) *Explosives*, revised and extended edn. Weinheim: VCH.

Dobratz BM and Crawford PC (1985) *LLNL Explosives Handbook*. CA: Lawrence Livermore National Laboratory.

Urbanski T (1964–1967) *Chemistry and Technology of Explosives*, 4 vols. Oxford: Pergamon Press.

Picatinny Arsenal (1960–1982) *Encyclopaedia of Explosives and Related Items*, vols 1–10. Dover, NJ: Picatinny Arsenal. (US AARADCOM).

F

FACIAL IDENTIFICATION

Contents

Computerized Facial Reconstruction
Facial Tissue Thickness in Facial Reconstruction
Lineup, Mugshot Search and Composite
Photo Image Identification
Skull-photo Superimposition

Computerized Facial Reconstruction

G Quatrehomme, Laboratory of Forensic Pathology and Forensic Anthropology, Faculty of Medicine, Nice, France
M Y İşcan, Adli Tıp Enstitüsü, Istanbul Üniversitesi, Istanbul, Turkey

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0765

Introduction

Identification of human remains has been a major problem for the medicolegal system. Thousands of people every year are buried with their identity and that of their killer unknown. Mass murders have been committed in many parts of the world, e.g. in Europe, Africa and South America, by dictatorial regimes or as a result of 'ethnic cleansing'. In many countries there is a lack of identification records, such as dental charts or DNA databases, or even ordinary identification cards. In affluent countries, not every individual has a dental record or has been reported to the police as missing. The only approach to identification may be facial reconstruction if the remains are entirely decomposed.

In general, facial identification techniques depend upon the available evidence and therefore may be classified as follows:

- *Video or photographic comparison* Comparison of a video image or photograph with the actual face, video or photograph of a suspect.
- *Skull–photo superimposition* Overlaying of a facial image (portrait, photograph or video footage) on a skull.
- *Facial restoration* When sufficient soft tissues persist on the head, the face may be restored to its original appearance.
- *Facial reconstruction* This technique is applied to the skull that has no persistent soft tissue.

The facial reconstruction technique has emerged as an important tool in the identification of unknown human remains. It differs from facial restoration, which deals with a face that is partly present but altered as a result of decomposition or trauma. The ultimate goal is to help identification of the unknown person. The reconstruction can be two-dimensional or three-dimensional, both of which can be performed artistically or done on a computer.

Obviously the relationship between every point of the bony face and the corresponding point on the soft tissues cannot be known with precision. Unfortunately the skull cannot give every clue needed to permit a perfect reconstruction. Although soft tissue thickness and bony variation and dimensions in many parts of the face are known for many populations, subtle details of the face will probably never be learned from the bony frame. Therefore, if the reconstructed face leads the investigator to the right person, or stimulates the public eye to recognize the picture, the mission may be considered accomplished. Despite these limitations, this form of identification is becoming an important tool for medicolegal investigation and is attracting research interest. Although there is no systematic survey or scientific test, the identification success rate is estimated to be about 50% or better.

In spite of many efforts to reconstruct the face from the skull, there still are very few research publications with scientific validation.

The aim of this article is to discuss the methods used in forensic facial reconstruction and the approaches taken to recreate the visual appearance of the unknown person. Over the years, faces and busts of historical, well-known people have been reconstructed: for exhibitions, for authentication (if a given skull belongs to a well-known personality), and to satisfy curiosity about the physical appearance of ancient people. The first attempts were made by German anatomists at the end of the nineteenth century in order to identify historically significant individuals. More recently, Russian anthropologists developed a manual for the three-dimensional technique and used it on ancient ethnic skulls as well as in forensic cases. In the last 25 years the two-dimensional technique has also gained in popularity.

Basis of Facial Reconstruction

Whichever technique is chosen, facial reconstruction requires a detailed analysis of the skull. The analysis should include the following:

- *Osteological examination* Skull morphology is assessed by observing osteological variation in terms of size and shape. In addition, anthropometric measurements are taken to discern the proportional relationship between the skull and the reconstructed face. Both morphological and metric examination also yield information about the age, sex, race and body size of the victim. The goal is to discern race, sex and age, as well as peculiarities of the skull and face such as asymmetry, antemortem health, pathology, trauma, cultural modification of the head and face, and individual habits. These can later be incorporated in the final reconstructed face. Standard university-based osteological courses are the best source for gaining an understanding of human variation of this nature.
- *Cephalometric/radiographic analysis* Although radiographs give some limited information about the biological characteristics mentioned above, it is also important for discovering peculiarities that are not otherwise visible to the naked eye. Both osteological and radiographic examinations should give enough information to create an image of the face which would 'individualize' the skull of a specific person.
- *Soft tissue thickness* Average soft tissue depth at specified skull landmarks is well known for many populations. Its variation has also been investigated for individuals who are thin, of average weight or obese.

Soft Tissue Thickness

One of the main issues in facial reconstruction is to understand the variation of soft tissue thickness in one's face, as well as within and between populations. The earliest research on facial tissue thickness goes back to the 1880s. It was carried out by inserting a thin knife blade at selected landmarks on cadavers. Later in the same century, measurements were taken with a needle. This needle technique was later adopted by the Japanese in the 1940s and the Americans in the 1980s.

Nevertheless, measurements on cadavers have been criticized for various reasons. The anthropological landmarks are not always easy to locate with certainty by palpating the face of cadavers, nor are the landmarks close enough to each other. Soft tissue deformation may occur during this crude process. The cadaveric alterations resulting from dehydration, rigidity and gravity may alter the soft tissue depth, probably even within a few hours after death. The bloating stage of the decomposition process increases the errors further. The correlation between the soft tissue depths measured on cadavers and living subjects is not known.

Recent investigators recommended that tissue depths should be measured *in vivo*. Commonly used equipment includes the ultrasonic echo detector, computerized tomography (CT) and magnetic resonance imaging. Furthermore, considering the diversity in human variation even in the same region, a large sample size is needed to assess the effect of such factors as age, sex and body weight. The dental occlusion pattern also plays a role in obtaining the right measurements. Using lateral cranial radiographs, it has been shown that the soft tissue depths showed age and sex differences as anticipated.

Two-dimensional Methods

Facial reconstruction rendering a two-dimensional (2D) view of a face can be accomplished by sketching on paper as well as on a computer. The easiest approach is to take a picture (full and/or profile view) of the skull at the Frankfort horizontal plane and enlarge it to its original size. Facial tissue thickness can be marked at known landmarks and these marked dots can be connected with each other to make a general outline of the face. Forensic artists are able to draw the face from this image, if properly guided. Therefore they usually work under the supervision of a forensic anthropologist who has already analyzed

the remains and its identifying skeletal characteristics, i.e. age, sex, race, body size, anomalies, trauma, pathological lesions, antemortem health status, and other unique features. The anthropologists should be able to describe where the eyes, ears, nose and mouth should be drawn in relation to the bony face. This requires a systematic understanding of human facial features. Based on these characteristics a face can be reconstructed, step by step. There are already several publications showing results of these studies. A similar attempt can be made using a lateral radiograph. Average soft tissue thicknesses are marked on the film or its copy and connected with each other.

Obviously certain aspects of the face must be modified, based on sex, race and age. The nose, for example, changes throughout life. Other features, such as the shape of the mouth and thickness of the lips, also change with age and their relationship with the facial skeleton. The 2D technique (sketches) has been used by some anthropologists to save time and reduce the cost of an artist.

Three-dimensional Methods

Facial reconstruction producing a three-dimensional (3D) result is the most desirable. It can be viewed in all aspects, rather than just a full face or profile view. The most traditional 3D approach is the manual build-up of the skull with a clay-like substance. Recently, however, computers have been used to give the 3D appearance.

Manual (plastic, sculptural) approach

The manual approach is one of the most popular 3D methods. Material (e.g. clay) is applied directly onto the skull, using the relationship between bone and soft tissues. This method was developed by German anatomists in the 1880s, and then adopted and modified by Russians, eastern Europeans and eventually the rest of the world. Its forensic use, however, goes back to the late 1940s. Today, practically every forensic artist, pathologist and dentist has experimented with the technique. There are even university courses designed to teach this subject.

When developing this 3D method, it is also necessary to assess skull morphology and metric characteristics, as was done for the two-dimensional analysis. In starting the reconstruction, marks are placed on precise anthropological points where the average tissue depths are known for the specific age and sex of that population. The space between these points is filled with clay or a similar substance (**Fig. 1**). Bit by bit the whole face is reconstructed (**Fig. 2**). Areas such as the ears, eyes, nose, mouth and lips are difficult to place because they do not have clear bony indicators. The literature varies as to the exact location of these facial structures.

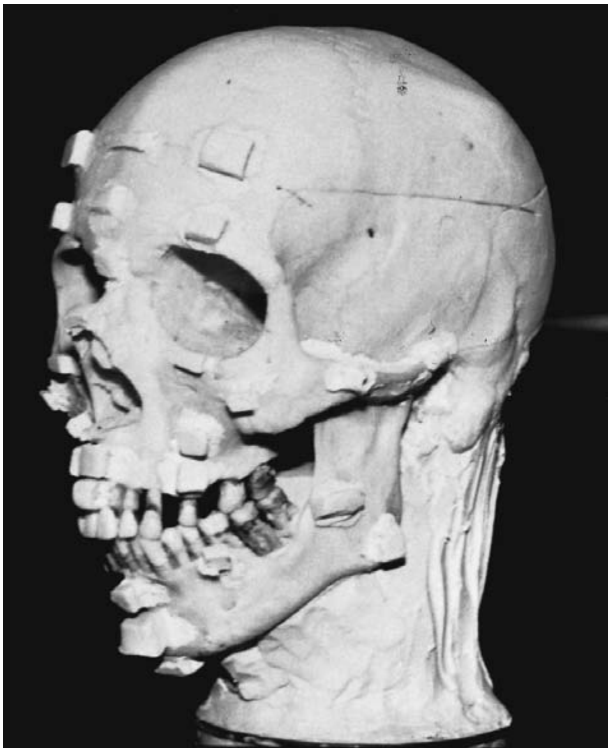


Figure 1 Classical manual 3D reconstruction: beginning of the reconstruction.

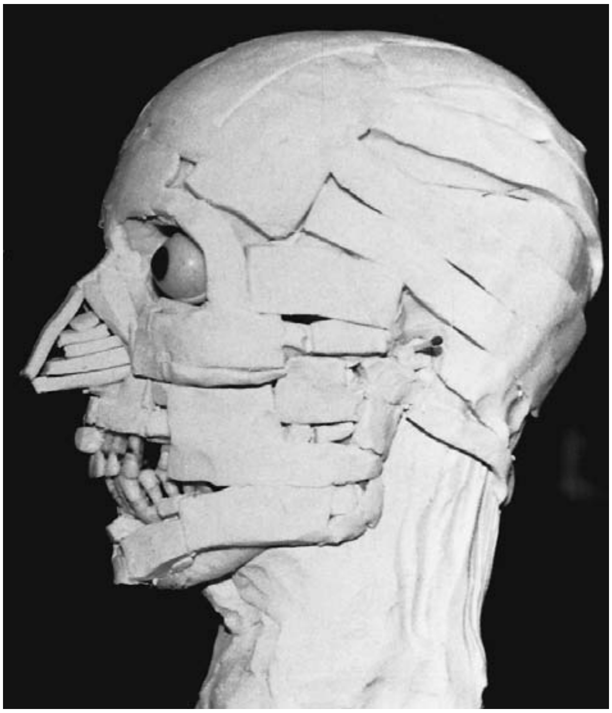


Figure 2 Classical manual 3D reconstruction: end of the reconstruction.

There are rarely experimental comparisons between the reconstructed face and the actual face of the deceased (**Fig. 3**). Similarly, there has not been any systematic comparison of the same face reconstructed in two and three dimensions. In one experiment, two anthropologists produced sketches (full and profile views) and clay images. When compared, it was noted that the similarities were slight. Yet the two images showed similar proportions in the orbit height, nose length and total, upper and lower face dimensions in the profile view.

Computerized methods

Usually computerized methods allow 2D rendering but some computers can generate a virtual image. Two-dimensional programs operate mostly by importing images or morphing. In one study using a virtual captured digitized image of the skull, the soft tissue contour was drawn using the soft tissue depths. The program provided an average of a series of features from a scanned face database. The average feature was then superimposed, like a mask upon the face. This method may limit the possible variation around the 'average' face. Newer programs have become more flexible and the process of reconstruction can be constantly checked by looking at a previous step.

The computerized 3D facial reconstruction has been developed by only a few scientists. Many are at an experimental stage. Successful identification should increase if a person is observed from different directions. Such data can be obtained by CT. Very often the conventional clinical CT protocol uses a plane which is parallel to the Frankfort horizontal. Laser scanning systems can also capture 3D images. In one study, for example, a color laser scanner is used to generate a representation of the scanned object as a 256×256 matrix. A wire frame of 256×256 radii is reconstructed. The wire frame matrix of the skull is transformed, using the tissue depth distances, through an algorithm which generates a 3D facial reconstruction.

The advantages of computerized methods are numerous because of the speed and the possibility of rapidly editing several versions of the reconstruction. Characteristics such as obese versus emaciated face can be easily altered. Color and textures can be changed quickly. The image can be rotated in three dimensions. The image could cycle through a variety of versions of the reconstruction based on the known range of tissue depth or other feature. There are

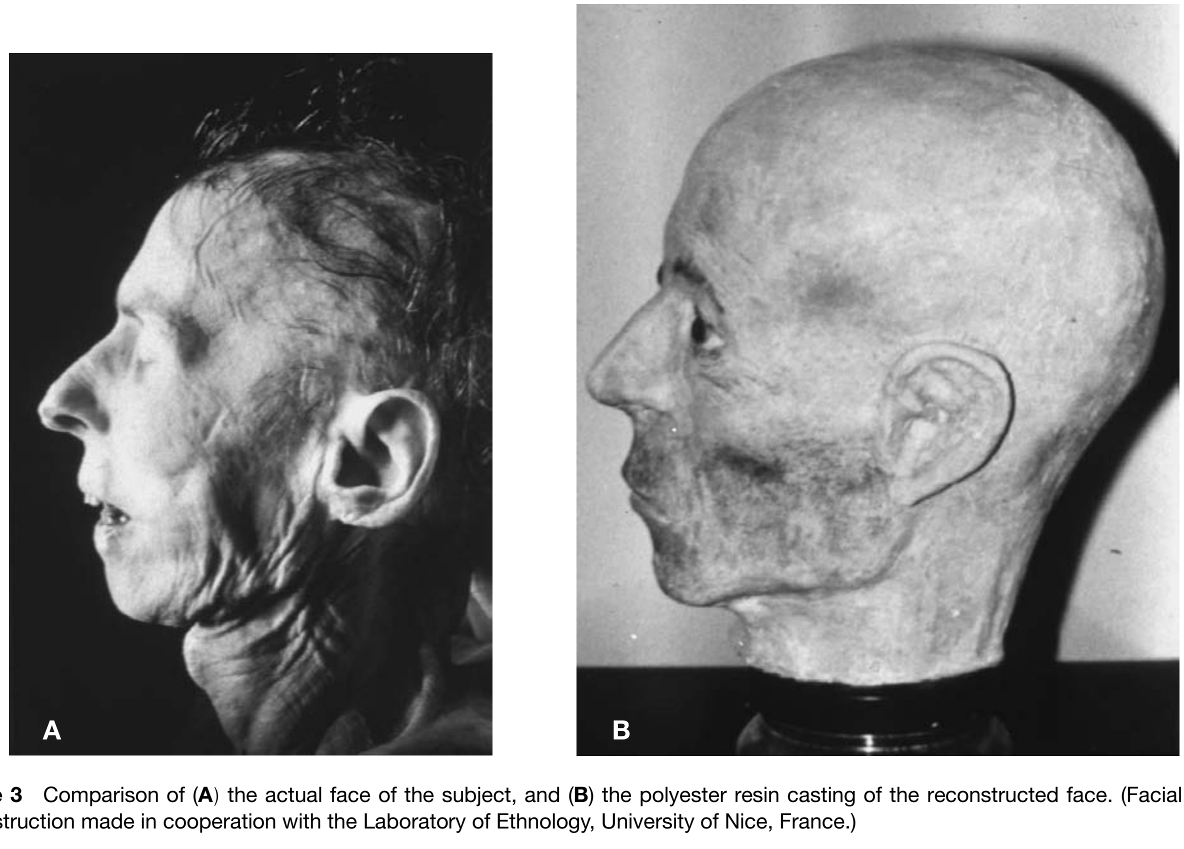


Figure 3 Comparison of (**A**) the actual face of the subject, and (**B**) the polyester resin casting of the reconstructed face. (Facial reconstruction made in cooperation with the Laboratory of Ethnology, University of Nice, France.)

additional benefits of a computerized reconstruction. Replication is possible when reconstruction is developed from one skull and uses the same computer setup. The image can then be animated to display a moving face. A 2D image can easily be written on a CD-ROM disc or recorded on a videotape. The image can be sent via the Internet or made accessible on the web. Age changes can be made for an older person. The face can be compared with the digitized image databases of missing persons. It is possible to rotate, zoom and alter the original image. The skull and virtual facial reconstruction can be superimposed on each other and the virtual facial reconstruction can be compared with the actual photograph of the missing person.

A review of the literature reveals that there are only a few articles on the subject of facial imaging and computer-assisted reconstruction. Only a few of the computed facial reconstruction programs are true 3D. One of them is based on the distortion of volumes assessed by the process of field morphing, using multiple pairs of disks. In this technique the original skull (S1) is altered into another (S2), and the same algorithm is applied to the original face (Actual Face AF1) to get the unknown face (Reconstructed Face RF2). The disks have to be placed on key levels all around the reference skull to be reconstructed. In another study two sets of models are used. The first set consists of a skull S1 and a known facial cast F1, the reference set. The second set is the skull to be reconstructed (S2), and the 'unknown' face (RF2) to be produced. Data acquisition was made by the digitalization of the two data sets of S1, S2 and F1 by CT. CT of a living subject is impossible in this protocol because very thin slices are needed and the amount of radiation received by the subject could be too much. CT of the head of a cadaver may raise ethical problems. As an alternative, the skull S1 and the mold of the face of the cadaver (F1) are used. CT obtained a set of slices with an interslice distance up to 1 mm. Those slices were stacked into volumetric images, then the skull and facial model were produced by an isosurface extraction algorithm. The method of reconstruction was a global parametric algorithm that transforms the reference skull S1 into S2. Then the same algorithm is applied to the reference head F1, to obtain the image of the unknown face (RF2). The experiment allows researchers to compare the reconstructed face (RF2) with that of the actual face (AF2). It should be noted that the transformation was not based on the traditional landmarks of the skull, but rather only on some salient lines called crest lines. Mathematically these lines are defined by differential invariants and correspond to lines of absolute maxima of the largest principal curvature. On the skull these lines correspond, for example, to the outlines of the mandible, the orbits, the cheekbones or the temples. In fact, when the crest lines are extracted from each skull, the algorithm that matches the set of crest lines of the first skull on to the set of crest lines of the second skull can be obtained. The algorithm is able to generate directly a 3D image of the face from the unknown skull. The image can be easily translated, rotated, zoomed, moved in a continual sequence and altered in different ways, including adding hair or spectacles. In addition, a set of landmarks can be selected and, if necessary, the distance between face and landmarks may be computed.

Discussion

Facial reconstruction is a difficult task. Some of the problem areas are the nutritional status and individual aging intensity as reflected in the face. Necessary information needed, such as the details of the nose, eye, ear, lips and chin, is very difficult to guess from the facial skeleton. Because of discrepancies between reconstructed faces and actual faces, some anthropologists have expressed pessimism about the effectiveness of facial reconstruction. However, there may yet be unidentified relationships between bone and soft tissue.

The nose is a good example for illustrating the difficulty in facial reconstruction because it is one of the most important features of the face used in identification. Unfortunately, the bony frame is limited and the tip is varied in shape and its appearance is very difficult to predict from the nasal and other surrounding bones. Furthermore, some parts of the bony frame of the nose (e.g. the nasal spine) are often broken away. The cartilaginous part is often absent, which dramatically increases the difficulties of the reconstruction. There are also population and sex differences in the human face. Facial morphology also changes throughout life. Many individuals may have a longitudinally oriented groove in the nose tip, which represents the overriding of the lateral crus of the alar cartilage over the lateral nasal cartilage, but such a structure is difficult to predict precisely. Another difficulty is how to determine whether a reconstructed face resembles the photograph of the victim. The identification process is very complex. 'Random items', like the ponderal status, pilosity and the color of the hair and eyes, are always difficult to guess because of continual changes in these areas as a result of aging, individual preference and disease.

There is a lack of scientific validation of the methods summarized above. One of the main problems raised by 3D reconstruction is the quality of the results and replication. Obviously, replication is necessary

for scientific reliability. Most of the reconstructions presented in the literature have been performed quite intuitively and have thus not been tested for their validity. Recognition does not always depend upon the general morphology of the face. In a forensic case, a close relative, when shown the reconstructed face, said it did not look like the victim because the neck looked 'too long', and it had a 'smile'. He stated that in reality the victim 'rarely smiled'. Also, the victim had 'dimples', which were not predicted from the skull morphology. In an experiment using the clay technique, it was found that 19 out of 22 cases resembled the victim's photograph. The most extensive attempt at validation was carried out by making duplicates of six skulls and letting two experienced artists perform the reconstruction. Then the two reconstructions were compared with the actual faces of the deceased individuals and with each other. The results showed that there was physiognomic similarity between the blind reconstructions, and the actual faces ranged from acceptable to excellent.

There are basically two scientific issues that need to be resolved in attempting to reconstruct a face from the skull of an unknown person. First, reliability refers to accuracy of the results and is an area in which some progress has been made. In several experimental cases, similar faces can be replicated by the same experimenter. In some cases, two different experimenters using the same methodology have obtained similar results. Second, validity refers to the truthfulness of the result, in which one expects to see a resemblance between the reconstruction and the victim. This aspect of facial reconstruction has been randomly attained and is inconsistent even with the same experimenter. In many cases it has been a 'hit and miss' result.

While forensic scientists are aware of these difficulties, there is always the question of the practical use of the results, that is, whether the reconstructed image can lead the police to the identity of the victim. There are some suggested solutions to these problems to assist the police in the identification process. One is to provide several shapes of one variant, like nose shape, ear projection, lip thickness, hair color and pilosity, ponderal status, and so on.

In conclusion, facial reconstruction has become an important and scientifically changing field of anthropology. Many techniques have been developed in the last 100 years to predict the facial appearance of an unknown person from his or her skull. In spite of the fact that there are many practitioners of this approach to identification, the relationship between bone and the corresponding soft tissue is not well known and knowledge of soft tissue depth itself is not sufficient to perform accurate facial reconstruction. The result is therefore only an approximation of the real face. The best one can hope for is to provide a stimulus which may direct the police to the identity of the missing person. This may in turn lead to identification when incorporated with other information collected about the victim. Furthermore, it is recommended that this technique should be carried out only after all others have failed. It is feared that once a 'wrong' face is presented to the public, the real identity may never be known. Nevertheless, if the victim's general characteristics (age, sex, race, height, antemortem health status and personal artifacts) are presented to the general public through the mass media, there is a greater chance of establishing the identity of the victim.

See also: **Anthropology**: Overview; Skeletal Analysis; Morphological Age Estimation; Sex Determination; Determination of Racial Affinity; Bone Pathology and Antemortem Trauma in Forensic Cases. **Identification/Individualization**: Overview and Meaning of ID.

Further Reading

Aulsebrook WA, İşcan MY, Slabbert J and Becker P (1995) Superimposition and reconstruction in forensic facial identification: a survey. *Forensic Science International* **75**(2–3):101–120.

Chen M, Jones MW and Townsend P (1996) Volume distortion and morphing using disk fields. *Computers and Graphics* **20**(4):567–575.

Evenhouse R, Rasmussen M and Sadler L (1992) Computer-aided forensic facial reconstruction. *Journal of Biocommunication* **19**(2):22–28.

Fedosyutkin BA and Nainys JV (1993) The relationship of skull morphology to facial features. In: İşcan MY and Helmer RP (eds) *Forensic Analysis of the Skull; Craniofacial Analysis, Reconstruction, and Identification*, pp. 199–213. New York: Wiley-Liss.

George RM (1987) The lateral craniographic method of facial reconstruction. *Journal of Forensic Sciences* **32**(2):1305–1330.

Gerasimov MM (1971) *The Face Finder*. Philadelphia: Lippincott.

Helmer RP, Röhricht S, Petersen D and Möhr F (1993) Assessment of reliability of facial reconstruction. In: İşcan MY and Helmer RP (eds) *Forensic Analysis of the Skull; Craniofacial Analysis, Reconstruction, and Identification*, pp. 229–246. New York: Wiley-Liss.

İşcan MY (ed) (1993) The introduction of techniques for photographic comparison: potential and problems. In: İşcan MY and Helmer RP (eds) *Forensic Analysis of the Skull: Craniofacial Analysis, Reconstruction, and Identification*, pp. 57–70. New York: Wiley-Liss.

İşcan MY and Helmer RP (eds) (1993) *Forensic Analysis of the Skull: Craniofacial Analysis, Reconstruction, and Identification*. New York: Wiley-Liss.

Krogman WM and İşcan MY (1986) *The Human Skeleton in Forensic Medicine*. Springfield, IL: Thomas.
Macho GA (1989) Descriptive morphological features of the nose: an assessment of their importance for plastic reconstruction. *Journal of Forensic Sciences* **34**(4):902–911.
Miyasaka S, Yoshino M, Imaizumi K and Seta S (1995) The computer-aided facial reconstruction system. *Forensic Science International* **74**(1–2):155–165.
Monga O and Benayoun S (1995) Using partial derivatives of 3D images to extract typical surface features. *Computer Vision and Image Understanding* **61**(2):171–189.
Nelson LA and Michael SD (1998) The application of volume deformation to three-dimensional facial reconstruction: a comparison with previous techniques. *Forensic Science International* **94**(3):167–181.
Prag J and Neave R (1997) *Making Faces: Using Forensic and Archaeological Evidence*. London: British Museum Press.
Quatrehomme G, Garidel Y, Grevin G, Liao Z, Bailet P and Ollier A (1995) Method for identifying putrefied corpses by facial casting. *Forensic Science International* **74**(1–2):115–124.
Quatrehomme G, Garidel Y, Grevin G, Liao Z, Boublenza A and Ollier A (1996) Facial casting as a method to help identify severely disfigured corpses. *Journal of Forensic Sciences* **41**(3):518–520.
Quatrehomme G, Cotin S, Subsol G *et al.* (1997) A fully three dimensional method for facial reconstruction based on deformable models. *Journal of Forensic Sciences* **42**(4):647–650.
Rhine JS (1984) Facial reconstruction in court. In: Rathbun TA and Buikstra JE (eds) *Human Identification: Case Studies in Forensic Anthropology*, pp. 357–362. Springfield, IL: Thomas.
Shahrom AW, Vanezis P, Chapman RC, Gonzales A, Blenkinsop C and Rossi ML (1996) Techniques in facial identification: computer-aided facial reconstruction using a laser scanner and video superimposition. *International Journal of Legal Medicine* **108**(4):194–200.
Snow CC, Gatliff BP and McWilliams KR (1970) Reconstruction of facial features from the skull: an evaluation of its usefulness in forensic anthropology. *American Journal of Physical Anthropology* **33**(2):221–228.
Subsol G, Thirion JP and Ayache N (1998) A general scheme for automatically building 3D morphometric anatomical atlases: application to a skull atlas. *Medical Image Analysis* **2**(1):37–60.
Szeliski R and Lavallée S (1994) Matching 3-D anatomical surfaces with non-rigid deformations using Octree-splines. *International Journal of Computer Vision (Electronic Version)* **2**:171–186.
Thirion JP and Gourdon A (1995) Computing the differential characteristics of isointensity surfaces. *Computer Vision and Image Understanding (Electronic Version)* **61**(2):190–202.
Tyrrell AJ, Evison MP, Chamberlain AT and Green MA (1997) Forensic three-dimensional facial reconstruction: historical review and contemporary developments. *Journal of Forensic Sciences* **42**(4): 653–661.

Facial Tissue Thickness in Facial Reconstruction

W A Aulsebrook, University of Natal, Durban, South Africa

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0768

Introduction

There are many instances of human remains where the only item available for identification is a denuded skull. How useful it would be, to be able to replace the missing soft tissues of the face and to see what the individual looked like in life. This is, in fact, the goal of forensic facial reconstruction (FFR), namely, to replace an artificially rendered version of the facial soft tissue over the unidentified skull, in such a way as to suggest the identity of the deceased. It has been suggested that the word (facial) reconstruction, be replaced by the term reproduction or approximation. For the purposes of this article the more widely used term, reconstruction, will be used.

In forensic science, facial soft-tissue depth measurements are of use in the following instances:

- Assisting in the alignment of a face image to a skull image in a technique called skull-photo superimposition. This is an estimation of 'fit' between two known images.
- Two-dimensional facial reconstruction where a soft-tissue profile is calculated, drawn and adapted to a tracing of the skull profile.
- Computerized application of a reconstructed or stored face image to a found skull.
- Three-dimensional rendering of the face, in clay or other malleable material.

The chief thrust of this article is to discuss the amount and form of facial soft tissue and the processes involved in three-dimensional FFR.

At the outset we should grasp the limits of FFR. What it is not able to do is to reproduce the exact features of an unknown face. What is possible, however, is that it can produce a semblance of the victim's appearance that could just be of assistance in the identification process.

Concept of Facial Reconstruction

How can one learn more about facial soft-tissue architecture? Certainly, open dissection is a good start. In fact, a practical understanding of head and neck anatomy should be a prerequisite for all those intending to work in FFR. The Further Reading list at

the end of this article refers to in-depth surveys of research work done in the field of FFR; comprehensive summaries of major works on forensic anthropology, and reviews on the current status of facial reconstruction.

The earliest tables of soft-tissue depth measurements were drawn up at the end of the nineteenth century (when the faces of notable people such as Kant, Schiller and Bach were reconstructed). Using thin knives and needles, investigators probed and recorded the facial soft-tissue depth of cadavers at a number of landmarks. Only in the third decade of the twentieth century was facial reconstruction used for forensic purposes. Over the years, tables of measurements were established for the three major ethnic groups: Caucasoid, Mongoloid and Negroid. The Negroid Group was composed of African-Americans. These are a blend of Negroid, Caucasoid and Mongoloid types. Although no ethnic group is, correctly speaking, biologically homogenous, there are certain isolated groups that have resisted crossgroup intermixture for long periods. In order to obtain a more homologous set of measurements for the Negroid, a table of facial soft-tissue depths for the African Zulu male was published in 1993. The data derived from the study are linked with the production of midline and oblique face profiles and are best understood in their entirety in the original work. As is the case in this last study, more researchers are now measuring tissue thickness using cephalometric radiography, ultrasonography and, more recently, computed tomography and magnetic resonance imaging. Various facial tissue depth measurements are also available for Caucasoid children, though these are representative of humans in their most rapidly changing stage of life.

In order to compare the measurements of the different studies, soft-tissue measurements should be taken at the same landmarks in each study. The most commonly used landmarks are illustrated in **Fig. 1**. Although the Mongoloid landmarks are fewer in number, they correspond to those used on white and black studies (see Table 2). There are also a number of smaller scattered studies on Caucasoid and Negroid groups, listed in the review mentioned in the Further Reading list.

Of course the value of any measurement is limited to a single locus and offers no information on the depth or character of the adjacent tissue. This means the person doing the modeling (the 'sculptor') begins work with a very limited amount of information indeed.

There is some disagreement as to who should carry out the practical work of FFR. It seems the ideal situation should occur when the attributes of scientist and artist exist in the same worker, but this is rare. Usually the anthropologist works together with an

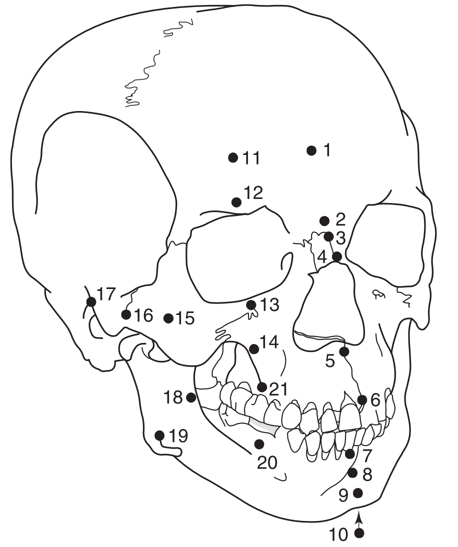


Figure 1 Landmarks for soft-tissue depth measurements. See Table 1 for landmark names corresponding to the numbers on the illustration.

artist to model the face. Some feel it unnecessary to use trained artists or sculptors, as the outcome of the work can be negatively affected by their creative tendency. This could well be the case if the artist is given free reign for interpretation. On the other hand, if 'trained' scientific experts are used, they need to have a fair modicum of manual dexterity and modeling skill to enable them to reconstruct a realistic-looking face. The ideal kind of worker is a professionally trained realist sculptor, such as a portrait artist, who can honestly render a three-dimensional face that looks like a face. In all cases, artistic ability should be respectful of, and subservient to, the dictates of science. In short, there is a distinct need for artistic talent, but not for artistic license.

There are fundamentally two different approaches to reconstructing the human face. The first makes use of objectivity, measurements and rules. The second, made popular by Russian workers, is based on anatomical knowledge, critical observation and subjective evaluation. The two methods could be called morphometric and morphoscopic, respectively. One might be forgiven for feeling that the morphoscopic method is unscientific and is therefore best left alone; however, there have been a number of talented workers who have used it with success. The morphoscopic method will be only briefly mentioned, as the main focus will be placed on the morphometric approach to facial reconstruction.

In order to make FFR a scientifically acceptable technique, an objective method must be devised that can produce the same result each time. Theoretically, if three sculptors are supplied with identical information and copies of a skull, they should be able, using a

standardized and controlled technique, to reconstruct three similar faces. If this can be accomplished, a reliable and repeatable method of facial reconstruction will have been produced, which will have the added advantage of being granted weight in the eyes of the law.

With this in mind the technique should conform to certain regulations. Firstly, only standard landmarks may be used. Secondly, only average soft-tissue measurements may be employed. Thirdly, the areas between landmarks must be filled in with mathematically graded clay strips and filler material. Finally, the sculptor must make use of 'rules of thumb' and controlled modeling techniques. Admittedly, the result may be a rather bland approximation that relies on the general form of the skull for any suggestion of likeness. The paradox is that people are usually remembered for their *differences* from the average; that is, for their peculiarities.

Practical Process

In all cases, the sculptor should be provided with as much information on the deceased as possible i.e. ethnic group, age, gender, stature and weight.

In most cases the mandible is attached to the skull in what is called centric occlusion: with upper and lower teeth together. Preferably, the mandible should be placed in 'centric relation', with a freeway space between the teeth. This represent a jaw that is relaxed. People do not normally walk around with their teeth clenched.

There are differing opinions as to whether the reconstruction should be carried out on the original skull or whether a duplicate casting should be used. If a casting is made, it is well to bear in mind that distortion can occur during the impression-taking process. Also, if the copy is cast in plaster of Paris a uniform dimensional change could take place during the setting phase. The thoroughly cleaned skull (or casting) is then mounted on a stand, ready for the modeling process.

Morphoscopic method

In this method, the face is considered as being individually highly characteristic. The skull is first examined for any 'clues' that could suggest the amount of muscle development, for example the size and shape of the mastoid process or the degree of eversion at the angle of the mandible. Reconstruction begins with modeling three of the main muscles of mastication: the temporalis, masseter and the buccinator (some will disagree that the last named is a muscle of mastication). Care should be taken not to exaggerate the bulk of these muscles. Though powerful, they are not all that thick. For example, a fair amount of the volume of tissue overlying the ramus is made up of fat and parotid gland, which gives the side of the jaw a characteristic shape. Likewise, the buccinator, though plump in children (as an aid to suckling) is not at all that thick in adults. The pterygoid muscles do not play a role in defining the form of the face.

Next, the circular muscles around the mouth and eyes are built up. These are relatively weak muscles and are quite thin and flat except for their junctions with other muscles, for example the commissures at the corners of the mouth. Other tissues to be added are the parotid glands and any fatty deposits, if suspected. Finally a layer of 'skin', varying in thickness between 0.5 and 1.0 cm, is adapted to the surface of the reconstruction and textured. The features are then modeled, more by anatomical knowledge, feel and experience, than by rules. Some sculptors use the morphoscopic approach throughout, but every now and again pierce the clay at the landmarks to check for average depth. The stages in the process are illustrated in **Fig. 2**.

Morphometric method

To begin with there must be some way of indicating the depth of 'tissue' to build up at the landmarks. Innovative workers have used clay mounds, rubber cylinders, carding strips and matchsticks as depth indicators. In the case of the reconstruction illustrated in **Fig. 3**, thin balsa wood rods were used.

The rods are cut to lengths matching the data in the appropriate tables of soft-tissue depths (**Tables 1** and **2**) and stuck to the skull (or cast) at the relevant landmarks. Small, 1.5 cm 'islands' of clay are built up to the height of each rod to prevent their accidental displacement. Flattened strips of clay, 1 cm wide and evenly graded in thickness from one depth rod to the next, are placed between the islands. In a similar manner, the voids between islands and strips are filled in, the rods removed, and the surface smoothed off, taking care not to thicken or compress the clay.

Except for the features, the whole skull and face are covered in this manner. The stages in this process are shown in Fig. 3. Drawn guides to the reconstruction of the features are illustrated in **Fig. 4**, and are explained in the following text.

Eyes Most sculptors insert artificial eyes into the sockets, although some prefer to construct eyeballs from clay, with the irises carved out. The plausible reason for this is that the color of the deceased's eyes is usually unknown. Perhaps, if the skull is that of a

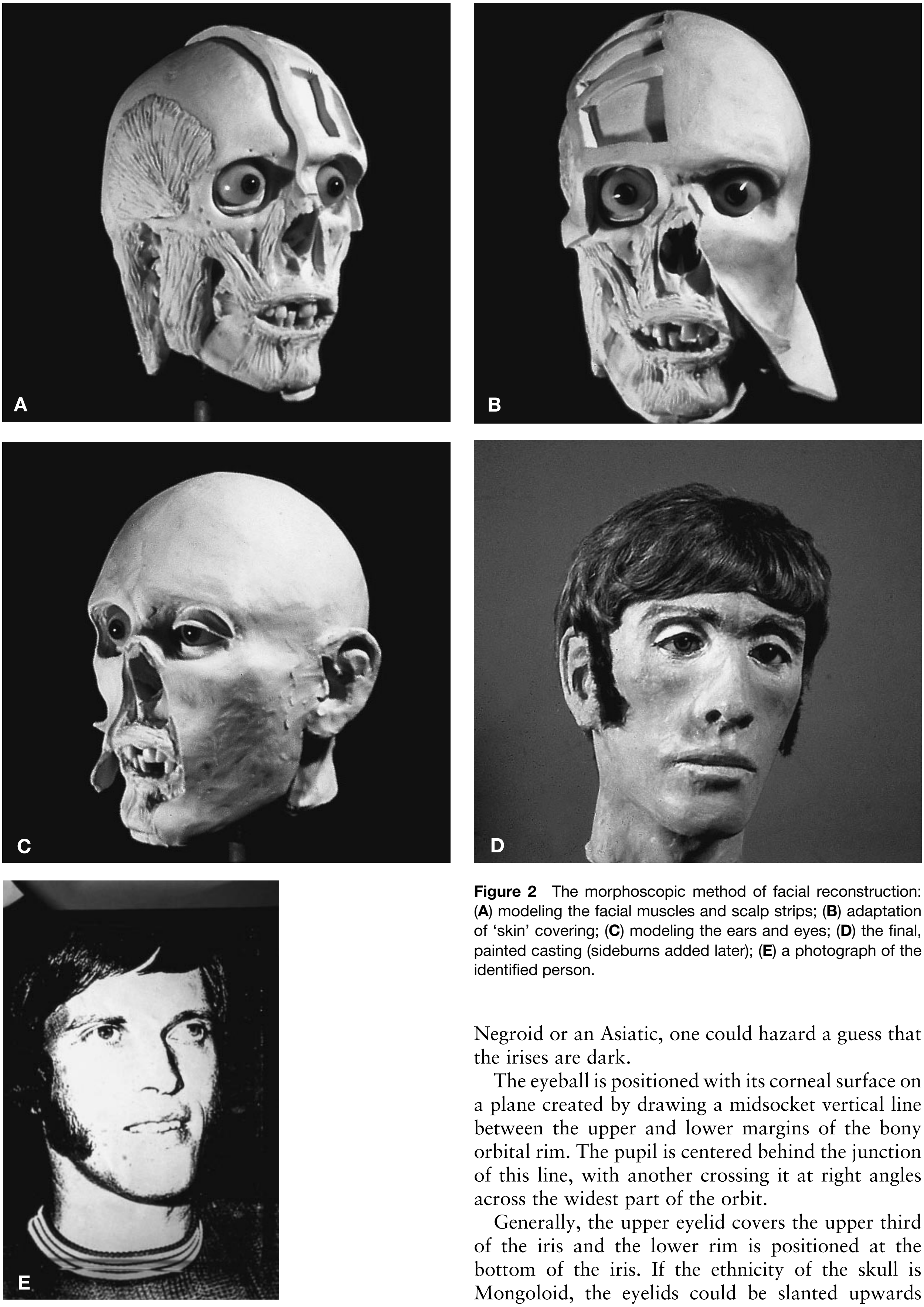


Figure 2 The morphoscopic method of facial reconstruction: (**A**) modeling the facial muscles and scalp strips; (**B**) adaptation of 'skin' covering; (**C**) modeling the ears and eyes; (**D**) the final, painted casting (sideburns added later); (**E**) a photograph of the identified person.

Negroid or an Asiatic, one could hazard a guess that the irises are dark.

The eyeball is positioned with its corneal surface on a plane created by drawing a midsocket vertical line between the upper and lower margins of the bony orbital rim. The pupil is centered behind the junction of this line, with another crossing it at right angles across the widest part of the orbit.

Generally, the upper eyelid covers the upper third of the iris and the lower rim is positioned at the bottom of the iris. If the ethnicity of the skull is Mongoloid, the eyelids could be slanted upwards

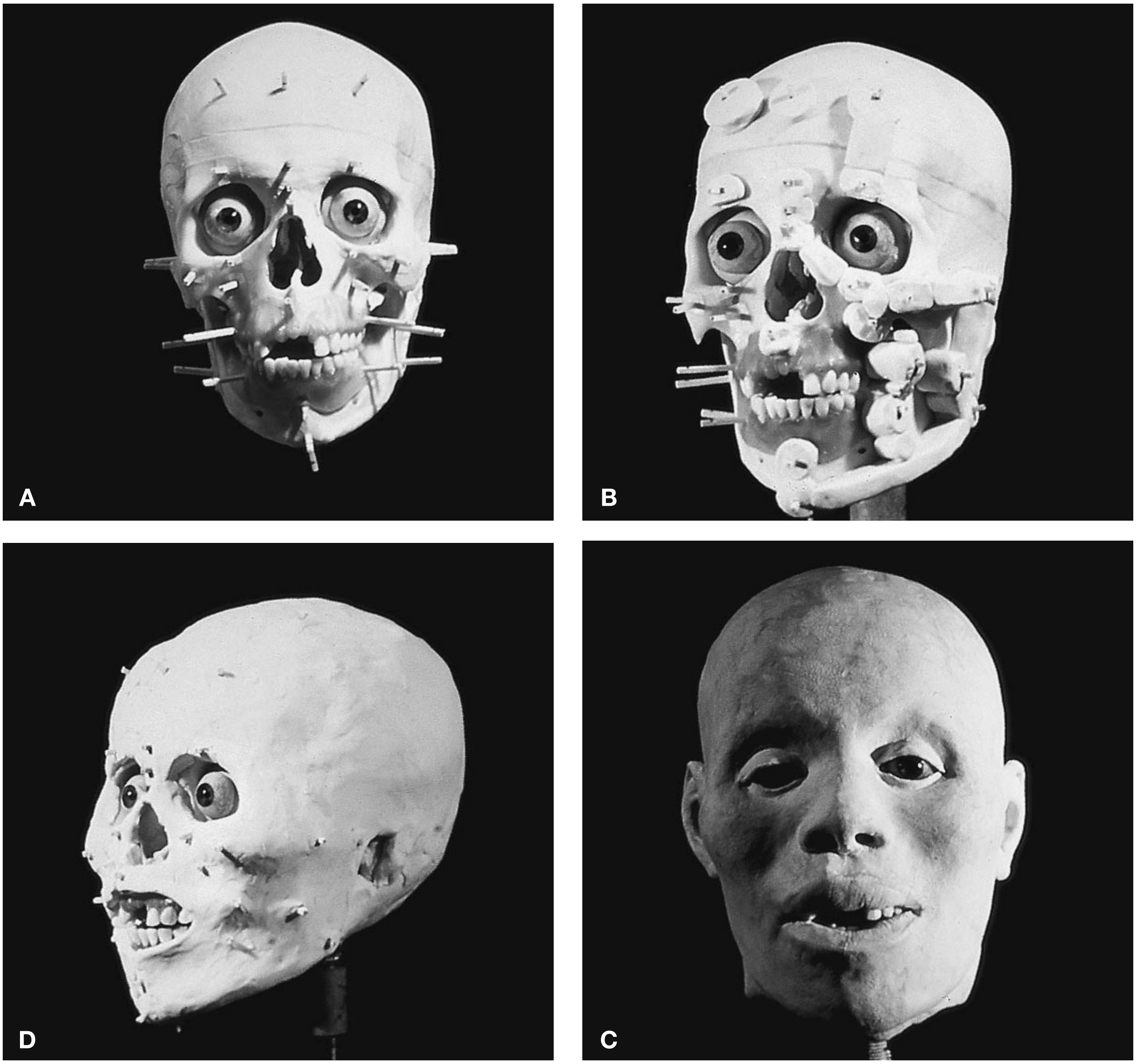


Figure 3 The morphometric method of facial reconstruction: (**A**) showing depth pegs stuck to skull, at landmarks; (**B**) islands and graded strips placed between pegs; (**C**) 'skin' surface smoothed off; (**D**) finished clay representation; (**E**) final painted casting.

and outwards. Otherwise, the slant of the eye can be roughly determined by a line running from the lachrymal opening in the medial wall of the orbit to the tubercle on the lateral orbital rim. In modeling the eyelids, one should take into account the thickness of the inner canthus (2 mm) and outer canthus (4 mm).

Nose The nose is considered a dominant feature. Yet, if it is within the bounds of average for that particular ethnic group, it is hardly noticed. Only if out of the ordinary is it of any significance in recognition. Except for the nasal bones, there is no underlying bony support for the soft tissues of the nose. Despite this, certain rules have been developed, the main ones being aimed at the length, height and width of the nose.

Table 1 Facial tissue thickness (mm) of American Whites

Landmark	*Emaciated*		*Normal*		*Obese*	
	Male (3)	*Female (3)*	*Male (37)*	*Female (8)*	*Male (8)*	*Female (3)*
Midline						
1. Supraglabella	2.25	2.50	4.25	3.50	5.50	4.25
2. Glabella	2.50	4.00	5.25	4.75	7.50	7.50
3. Nasion	4.25	5.25	6.50	5.50	7.50	7.00
4. End of nasals	2.50	2.25	3.00	2.75	3.50	4.25
5. Midphiltrum	6.25	5.00	10.00	8.50	11.00	9.00
6. Upper lip margin	9.75	6.25	9.75	9.00	11.00	11.00
7. Lower lip margin	9.50	8.50	11.00	10.00	12.75	12.25
8. Chin-lip fold	8.75	9.25	10.75	9.50	12.25	13.75
9. Mental eminence	7.00	8.50	11.25	10.00	14.00	14.25
10. Beneath chin	4.50	3.75	7.25	5.75	10.75	9.00
Bilateral						
11. Frontal eminence	3.00	2.75	4.25	3.50	5.50	5.00
12. Supraorbital	6.25	5.25	8.25	7.00	10.25	10.00
13. Suborbital	2.75	4.00	5.75	6.00	8.25	8.50
14. Inferior malar	8.50	7.00	13.25	12.75	15.25	14.00
15. Lateral orbit	5.00	6.00	10.00	10.75	13.75	14.75
16. Midzygomatic arch	3.00	3.50	7.25	7.50	11.75	13.00
17. Supraglenoid	4.25	4.25	8.50	8.00	11.25	10.50
18. Occlusal line	12.00	11.00	18.25	17.00	23.50	20.25
19. Gonion	4.50	5.00	11.50	12.00	17.50	17.50
20. Sub M2	10.0	9.50	16.00	15.50	19.75	18.75
21. Supra M2	12.00	12.00	19.50	19.25	25.00	23.75

Adapted with permission from Rhine J S and Moore C E (1982) *Facial Reproduction Tables of Facial Tissue Thicknesses of American Caucasoids in Forensic Anthropology*. Maxwell Museum Technical Series #1, Albuquerque, NM.

The nose length and height are determined by drawing a line tangent to the terminal third of the nasal bones in the midline and continuing it downward and forward to meet another line, extending the axis of the anterior nasal spine, running upwards and outwards.

There are various rules for the width of the nose:

- It is the same as the inner canthal distance (which in turn is often claimed to be the same as the palpebral fissure length),
- It is 10 mm wider than the bony width in whites, and 16 mm in blacks.
- It is 1.67 times the width of the bony nasal aperture.
- It is one-third larger than the bony width (the pyriform aperture is three-fifths of the interalar width).
- A well controlled study of 197 skulls and face masks in the Terry collection showed the interalar width to be the pyriform aperture width (PAW) + 12.2 mm in Caucasoid males, and PAW + 16.8 mm in Negroid males.

Measurements of this sort need to be extended over a large number of samples, varying in gender, age and ethnicity, before they are of sound value.

The most characteristic shapes of the nose are the bridge curve (arched, straight, dipped, wide or narrow), much of which is suggested by the form and direction of the nasal bones, and the nose tip which has no bony clue and is highly variable. The alae differ more in size than they do in shape. The nostril shape is another variant. In the high narrow nose the oval-shaped nasal opening is angled more vertically, while in the wide flat nose it is horizontally positioned.

Mouth The mouth is a capricious feature, although in some ethnic groups it is rather characteristic of that group. For example, Negroid lips are almost inevitably everted and are thicker than Caucasoid lips. A rule-of-thumb is that the width of the mouth, or intercommissural distance, is equal to the interpupillary distance. Dentists are taught that the corners of the mouth are positioned at the rear edge of the upper canines, with the upper lip covering two-thirds of the upper anterior teeth. In some cases the upper lip reaches down to the incisal edges. The fullness and contour of the lips are dependent on ethnicity, age and dental support. The philtrum or vertical midline gully in the upper lip is also a variable structure. If the upper dental arch is wide, the

Table 2 Facial tissue thickness (mm) of American Blacks and (Negroids) and Japanese (Mongoloids)

Landmark	*Black*		*Japanese*		
	Male	*Female*	*Point*	*Male*	*Female*
Midline					
1. Supraglabella	4.75	4.50	m	3.00	2.00
2. Glabella	6.25	6.25	gl	3.80	3.20
3. Nasion	6.00	5.75	n	4.10	3.40
4. End of nasals	3.75	3.75	rhi	2.20	1.60
5. Midphiltrum	12.25	11.25	–	–	–
6. Upper lip margin	14.00	13.00	–	–	–
7. Lower lip margin	15.00	15.50	–	–	–
8. Chin-lip fold	12.00	12.00	ml	10.50	8.50
9. Mental eminence	12.25	12.25	pg	6.20	5.30
10. Beneath chin	8.00	7.75	gn	4.80	2.80
Lateral					
11. Frontal eminence (L)	8.25	8.00	–	–	–
Frontal eminence (R)	8.75	8.00	–	–	–
12. Supraorbital (L)	4.75	4.50	–	–	–
Supraorbital (R)	4.75	4.50	sc	4.50	3.60
13. Suborbital (L)	7.50	8.50	–	–	–
Suborbital (R)	7.75	8.50	or	3.70	3.00
14. Inferior malar (L)	16.25	17.25	–	–	–
Inferior malar (R)	17.00	17.25	–	–	–
15. Lateral orbit (L)	13.00	14.25	–	–	–
Lateral orbit (R)	13.25	12.75	ma	5.40	4.70
16. Zygomatic arch (L)	8.75	9.25	–	–	–
Zygomatic arch (R)	8.50	9.00	zy	4.40	2.90
17. Supraglenoid (L)	11.75	12.00	–	–	–
Supraglenoid (R)	11.75	12.25	–	–	–
18. Occlusal line (L)	19.50	18.25	–	–	–
Occlusal line (R)	19.00	19.25	–	–	–
19. Gonion (L)	14.25	14.25	–	–	–
Gonion (R)	14.75	14.25	go	6.80	4.00
20. Sub M2 (L)	15.75	16.75	–	–	–
Sub M2 (R)	16.50	17.25	m_1	14.50	12.30
21. Supra M2 (L)	22.25	20.75	–	–	–
Supra M2 (R)	22.00	21.25	m_1	10.20	9.70
Sample size	44	15		9	7
Average size	38.0	32.8			Over 40

[a] Adapted with permission from Rhine J S and Campbell H R (1980) Thickness of facial tissues in American Blacks. *Journal of Forensic Sciences* **35**: 960–963.
[b] Adapted with permission from Suzuki K (1948) On the thickness of soft parts of the Japanese face. *Journal of the Anthropological Society of Nippon* **60**: 7–11.

philtrum is frequently flat. The converse is also often true, especially if the nose is narrow and tipped upwards. The anteroposterior relationship of upper to lower lip is characteristic in cases of an open bite and a jutting lower jaw.

Ears As with the nose, ears are really only noticed when they are out of the ordinary: bat-eared, flat-eared, distorted, extra-large, extra-small, or where the bottom half of the ear juts out instead of the top. The ear is claimed by some to be the length of the nose and canted at 15° behind the vertical. Another suggestion is that the inclination of the ear is parallel to the forehead. The tragus of the ear is positioned over the bony external aperture.

Chin The soft-tissue chin is not merely an echo of the underlying bony chin. Generally, it is squarer and more robust in males than females.

Jowls The amount of fatty tissue on the face is variable and is distributed in the jowl area and neck. Not always cost-effective is the production of three facial reconstructions: lean, medium and fat. The measurements in Table 1 have taken this into account.

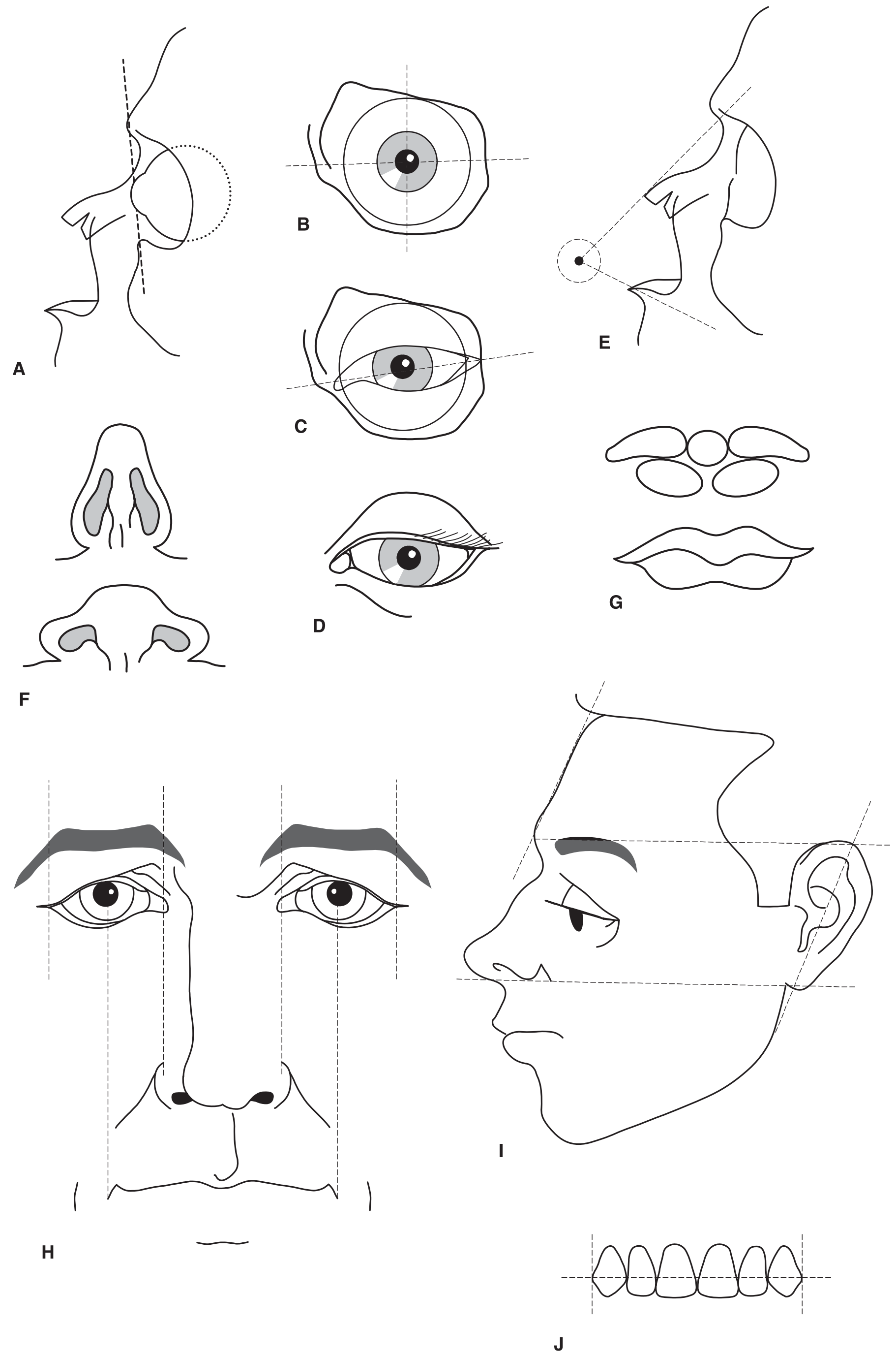


Figure 4 'Rules-of-thumb' used to guide reconstruction of facial features: (**A**) anteroposterior placement of eyeball; (**B**) centering the eyeball; (**C**) angulation of palpebral fissure; (**D**) eyelids in relation to the iris; (**E**) finding the nose tip; (**F**) nostrils from below; (**G**) division of lip into five masses; (**H**) alignment between eyes, nose and mouth; (**I**) length of ear and its angulation; (**J**) lip placement over upper teeth. Refer to the text for further explanation.

Finishing off and casting

The surface of the face is smoothed and textured to resemble human skin. In order to preserve the reconstruction and make it durable for presentation, a mould is fabricated of the whole head, and a casting made in hardened plaster of Paris. Leaving the face unpainted results in an undesirable ghostly appearance. Instead, the surface of the cast is painted in flesh tones. In the absence of other clues, it is necessary to guess the color of the skin. The use of 'add-ons', such as facial and head hair, glasses and earrings, must be avoided at this stage. A balding head or the wrong hairstyle can confuse identity. It is preferable to have

'add-ons' available for placement, if requested by the identifying witness.

Discussion

It seems sensible that soft-tissue measurements be recorded on a fleshed-out living subject in the upright position, rather than on a distorted supine corpse. Also soft-tissue depths are often derived from samples of subjects ranging across age and ethnic groups. Unless samples are kept within defined and controlled parameters they are little more than a measure of neutrality. For example, average measurements on a mixed-ethnic subject iron out the differences between the primary groups and provide tissue thickness lying midway between one group and the other. In living people, this is not necessarily found to be the case. It appears the mixed individual can inherit different features from each parent. If an examination of the skull reveals an index or feature of a particular ethnic group, then the measurements and anatomical form characteristic of that group should be used for constructing the inherited feature.

There was a time when FFR was viewed with suspicion and criticism. More recently, respectability has been improved through studies that attempted to evaluate the efficacy of FFR, with varying degrees of success. Some researchers have claimed as much as a 72% success rate with the morphometric methods of reconstruction. What is meant, when a reconstruction is claimed as 'successful'? Was it a success because it matched the deceased's face? Was it a success because it jogged the memory of an identifying witness? Or was it a success because it played a part in excluding other suspects?

In the case illustrated in Fig. 2, the 'success' occurred when a policeman who saw a photograph of the reconstructed face in the newspaper remembered having seen it somewhere before. He unearthed a 17-year-old newspaper cutting of the missing person and the final identification was confirmed after photo-superimposition of the teeth. The wig had been present from the start (not advisable, as already mentioned), but the sideburns were added after the victim had been identified.

It is unfair, at this stage of its development, to expect FFR to produce an accurate positive identification. It has value in a positive sense when it stimulates direct investigative leads to the unidentified person, and in a negative sense when it eliminates those suspects whose face could not possibly 'fit' the unidentified skull. At all times it should be understood that facial reconstruction is still only an adjunct in the identification process. The final identification of the victim must depend on other tried and proven methods of confirmation, such as dental identification and accurate DNA analysis.

The Future

What technology and new ideas are available for research in the field of FFR? Tools such as computer tomography, magnetic resonance imaging, photogrammetry and laser-scanning present us with masses of coordinate data on hard- and soft-tissue form. The challenging concepts of morphometrics (the fusion of geometry with biologic homology), finite elements analysis, Fourier shape analysis, neural nets and strange attractors, suggest different ways of looking at that data. The focus is swinging away from singularities (such as individual depth measurements) and is heading in the direction of continua. Even in the simplified morphometric technique (and for that matter the morphoscopic method), small but significant advances in the progress towards continuous outlines occurred though the use of drawn outlines for the lateral (and in one study, the lateral and oblique) profile of the face. The recent popularity of fuzzy logic is tempting us to discard 'either/or' concepts and to search for gradations of values between extremes.

The use of laser scanning to record the surface of the face, coupled with parallel processing and new Windows NT-based software, enables researchers, guided by average tissue depths, to adapt a stored face to an unknown skull and to edit and manipulate the face for a finer fit. Although speed, objectivity and the ease of circulating multiple copies of the finished result give it a certain edge over hand-sculpted efforts, there still remains the problem of using average measurements.

Sharing of visual image data between researchers in institutions and other computer stations is becoming available. A central workstation in a university computer laboratory can utilize advanced technology to produce a three-dimensional reconstruction, which can be instantly transmitted to a low-tech endpoint, for example a police station. The programs and the large amounts of data needed for such work are processed, manipulated and stored at the central terminal.

It is certainly impressive to be able to reproduce a face from masses of coordinate data. However, much of that data may well be redundant as far as an understanding of the relationship between hard and soft tissue is concerned. What should be investigated is the relationship that exists between the hard and soft tissues, which in turn suggests the formulation of predictive rules between the two. Put in another way, what is being suggested is to isolate smaller components of the skull and represent them in a form that can be statistically handled. They can then be compared with similar packages in the adjacent soft tissue

and, if there is any correlation, a richer set of rules for constructing the individual features could be formulated. A sort of: 'If I see *this* configuration on the skull, can I expect to see *that* type of form in the soft tissue?' Despite the doubts of some, this may need less information than we think. As a precursor to this, clusters of similar face types could be mapped, investigated for common forms, and average measurements recorded for each cluster. Of course, right now, the human mind is more adept at noting any of these facial categories and their significance than is the computer.

Despite these forthcoming innovations, present-day clay-based forensic facial reconstructions, utilizing average soft-tissue measurements, need not be abandoned. At the very least, and for some time, it will have a role to play in areas where sophisticated technology is not freely available. And speaking of sophistication and the future, it is daring but not absurd to postulate that the rapidly developing science of DNA mapping could perhaps one day locate the genes for facial appearance.

See also: **Facial Identification**: Lineup, Mugshot Search and Composite. **Anthropology**: Morphological Age Estimation, Determination of Racial Affinity; Sex Determination; Photo Image Identification; Computerized Facial Reconstruction; Skull-photo Superimposition.

Further Reading

Aulsebrook WA (1993) *The Establishment of Soft Tissue Thicknesses and Profiles for Reconstruction of the Adult Zulu Face*. PhD thesis, Johannesburg: University of the Witwatersrand.

Aulsebrook WA, Bekker PJ and İşcan MY (1996) Facial soft tissue thicknesses in the adult male Zulu. *Forensic Science International* 79(2):83–102.

Aulsebrook WA, İşcan MY, Slabbert JH and Bekker P (1995) Superimposition and reconstruction in forensic facial identification: a survey. *Forensic Science International* 75:101–120.

Berkovitz BKB and Moxham BJ (1988) *A Textbook of Head and Neck Anatomy*. London: Wolfe.

Caldwell PG (1986) New questions (and some answers) on the facial reproduction techniques. In: Reichs KJ (ed.) *Forensic Osteology*, pp. 229–254. Springfield, IL: Thomas.

Clement JG and Ranson DL (eds) (1998) *Craniofacial Identification in Forensic Medicine*. London: Arnold.

Farkas LG (1981) *Anthropometry of the Head and Face in Medicine*. New York: Elsevier.

Gerasimow MM (1971) *The Face Finder*. Philadelphia: Lippincott.

İşcan MY (1988) Rise of forensic anthropology. *Yearbook of Physical Anthropology* **31**:203–230. New York: Liss

İşcan MY and Helmer RP (1993) *Forensic Analysis of the Skull: Craniofacial Analysis, Reconstruction and Identification*. New York: Wiley-Liss.

Krogman WM and İşcan MY (1986) *The Human Skeleton in Forensic Medicine*. IL: Thomas.

Powell N and Humphreys B (1984) *Proportions of the Esthetic Face*. New York: Thieme-Stratten.

Prag J and Neave RAH (1987) *Making Faces: Using Forensic and Archaeological Evidence*. Texas A and M University .

Rathbun TA (1984) Personal identification: facial reproductions. In: Rathbun TA and Buikstra JE (eds) *Forensic Anthropology*, pp. 347–356. Springfield, IL: Thomas.

Stewart TD (1979) *Essentials of Forensic Anthropology*. Springfield, IL: Thomas.

Tyrell A, Rega E, Chamberlain AT and Green MA (1994) *Report: Facial reconstruction from the skull: a review of current techniques and results*. Department of Archaeology and Prehistory and the Department of Forensic Pathology, University of Sheffield.

Lineup, Mugshot Search and Composite

A M Levi, Israel Police Headquarters, Jerusalem, Israel

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0512

Introduction

Witnesses are often critical for investigations and trial evidence, simply because other forensic evidence is missing. However, people are not cameras. They cannot be treated, or considered, as physical evidence. The true value of the eyewitness is much more fragile, and we must handle his or her testimony with special care. One respected source estimates that 56% of false incriminations have been due to eyewitness error.

When witnesses personally know the culprit, facial identification methods are of course not needed. The witness simply reports that 'John did it'. Identification issues arise when the culprit is a stranger. The problem is then far more difficult, because the witness has usually seen the culprit for only a short time. Memory of the face is imperfect.

The police, who deal with facial identification, have three tools with which to utilize eyewitnesses. If they have a suspect, they conduct a line-up (**Fig. 1**). If they do not, they conduct a mugshot search (**Fig. 2**).

Figure 1 Six-person photographic line-up. The suspect is in the middle of the top row.

If that fails, they can fall back on the composite (**Fig. 3**).

The traditional police line-up has the witness view five to eight people, or up to 12 photographs, one of whom is the suspect. If the witness 'identifies' the suspect as the culprit, the suspect will usually be charged for the crime and convicted. The court requires no other evidence. If the witness chooses someone else, a foil, the police will usually release the suspect.

The mugshot search has the witness view photographs of suspected criminals, usually of people convicted for offenses. If the witness 'identifies' someone, that person becomes a suspect. In contrast to the line-up, however, that person does not necessarily become a defendant. The criminal justice system understands that witnesses sometimes mistakenly identify someone. In the line-up, such error will usually involve choice of a foil, a known innocent. The innocent suspect is thus protected somewhat from such mistakes. On the other hand, there are no known innocents among the mugshots. Anyone chosen becomes a suspect, and therefore there is no protection against mistaken identifications. The mugshot search is sometimes confused with the line-up. Police often fail to distinguish the two, presenting as evidence of identification a choice in an 'all suspect line-up' (a mugshot search). Courts often fail to note the difference.

The mugshot search, however, is still important. It sometimes provides a suspect, an essential ingredient for continued investigation. Once a witness has identified a suspect in this way, however, conducting a

Figure 2 Two mugshots.

line-up is considered inappropriate. The witness may choose the suspect based on his or her memory of the mugshot, rather than the culprit at the scene of the crime.

This is just one example of the fragility of eyewitness evidence. People are better in noting that a person or photograph is familiar than they are at remembering the circumstances in which they saw that person. There have been cases in which a witness has identified a person as the culprit (he or she looked familiar), but had actually seen the person in a perfectly respectable setting.

Since mental images of the culprit may be quite vague, witnesses can be susceptible to subtle and even unintended influence by police officers who have a hunch regarding the identity of the culprit. Good practice calls for the maximum safeguards, including having the identification procedure conducted by an officer who is not otherwise involved in the investigation.

The composite is the tool of last resort. A trained police officer may create a picture of the culprit's face with the help of the witness. Usually the police use a collection of examples of each facial feature (hair, nose, eyes etc.). Computer programs that add additional aids, such as being able to modify each feature, are being used frequently.

In contrast to the mugshot search 'identification', the composite is but a first step in finding a suspect. The composite is not a photograph, only a likeness – at best. A suspect is found only when some informant 'recognizes' the composite as a particular person.

The composite is no less immune than the other two methods from being a source of undue influence on the witness. Research has demonstrated that the constructer of the composite can influence the composite produced. The image of this composite can then interfere with the ability of the witness successfully to identify the culprit.

We will discuss the strengths and weaknesses of these identification tools, and explore potential improvements.

The Line-up

The line-up, being the major means of facial identification in court, is the chief source of false incriminations. This is caused by a serious gap between its actual and believed probative value. 'Identification'

Figure 3 Composite of one member of the photographic line-up (Fig. 1). (Which one? Answer at the end of the article.)

means ‘establishing absolute sameness,’ but in fact all witnesses do is *choose* a line-up member. Their mental image of the culprit cannot be perfectly identical even with a guilty suspect. Even if the mind were a camera, the culprit would have a different facial expression, be wearing different clothes, and so on. Witnesses compare line-up members with their mental image, and decide to choose when the similarity is great enough for them.

They need not choose anyone. Good police practice reminds them of this at the outset. However, in experiments when the suspect is innocent, witnesses choose someone 60% of the time. They are simply choosing the line-up member who is most similar to their mental image. An innocent suspect has as much a chance as any other line-up member as being the most similar, and therefore to be chosen. The probability of choosing the suspect is therefore $1/n$, where n is the number of line-up members. When $n = 6$, the innocent suspect will be chosen $60/6 = 10\%$ of the time.

Witnesses are usually confident in their ‘identification’, and the criminal justice system overbelieves confident witnesses. Thus, courts will usually try and convict these innocent suspects. Moreover, the percentage of innocent defendants who have been chosen in a line-up is far greater than 10%. We must also consider how often witnesses choose guilty suspects in a line-up, as only they will usually be tried. In experiments, only 44% of guilty suspects are chosen, and eyewitness conditions in the real world are far worse than experimental conditions.

For example, experiments have shown that witnesses are less likely to identify the culprit if a weapon was visible during the crime, and if the witness was the victim rather than a bystander. The underlying factor seems to be distraction from attending to the culprit’s face. When a weapon is present witnesses focus on it, and when they are the victim they are preoccupied with their fear and/or how to escape. Yet, only 8% of experimental conditions have the culprit wield a visible weapon, compared to 44% of real cases. The witness was the victim in only 4% of the experimental conditions, compared to 60% of real life. Furthermore, real-world distraction is much greater: the gun is pointed at the witness/victim, who is generally assaulted physically. Thus, an estimate of 30% identifications of the guilty in real cases may be too high.

Bayesian analysis shows that under these conditions (10% innocent defendants, 30% guilty ones), 25% of the defendants are innocent. This is true if, without the line-up, there is a 50:50 chance of the defendant being guilty. Even when, without the line-up, the prosecution has evidence leading to an 0.85 probability of guilt (a strong case indeed, without other forensic evidence), then 5.5% of defendants are innocent. Guilty beyond reasonable doubt? A scientist would hesitate to publish a finding with a probability of 0.055 of being wrong. Yet defendants are often sent to jail based on weaker evidence.

This analysis is based on the assumption of a perfectly fair line-up. Unfortunately, line-ups are sometimes not fair. Informing witnesses that they need not choose someone is not universal. Without that instruction, more than 60% of witnesses will choose, increasing the rate of mistaken identifications.

Furthermore, a line-up is only fair if the suspect does not stand out in any way, and if each member is similar enough to the suspect. Chances of choosing a suspect increase if he is the only unshaven member, for example, or if the culprit has black hair but some line-up members are blonde. If, for example, two foils in a six-person line-up have the wrong hair color, the actual size of the line-up for the witness is only four members. The witness can immediately discount those two foils. In addition, the witness of course must be given no hint, by design or unintentionally, as to who the suspect is.

The officers investigating the case, the same ones

who have determined who the suspect is, often conduct the line-up. Their belief in the suspect's culpability sometimes leads them to 'help' the witness.

Sometimes officers have honest intentions but err. They may have great difficulty rounding up enough appropriate foils. They may put more than one suspect in the same line-up, which also increases the likelihood of choosing an innocent suspect. More and more, photographic line-ups are replacing live line-ups, to cope with increasing logistical difficulties in recruiting foils. Yet photographs reduce the chances of identifying guilty suspects, and thus also increase the percentage of innocent defendants.

Even such a seemingly simple idea as choosing line-up members similar 'enough' to the suspect is controversial. It has been argued cogently that no logical means exists for determining how much is 'enough'. Too much similarity (twins as an extreme example) could make the task too difficult, again decreasing identifications of guilty suspects.

The alternative suggested is to choose foils who fit the verbal description the witness gave of the culprit. Since these descriptions tend to be rather vague, line-up members would not be too similar, but similar enough to prevent the witness from discounting foils too easily (as in the case of a blonde foil with a black-haired suspect).

A number of solutions have been suggested to decrease the danger of the line-up. The English Devlin Report recommended that no one be convicted on the basis of line-up identification alone. This suggestion has since been honored in the breach. In addition, we have noted that Bayesian analysis indicates that quite strong additional evidence is required for conviction beyond reasonable doubt. Courts often accept quite weak additional circumstantial evidence to convict.

The police can prevent the worst abuses of the line-up; for example, in addition to proper instructions to the witness, an officer not involved in the case should conduct the line-up. However, we have noted that even perfectly fair line-ups are quite dangerous.

Psychologists can testify in court as expert witnesses. The goal is not simply to decrease belief in the line-up; that would increase the number of culprits set free. Rather, the goal is to provide information that could help the court differentiate between the reliable and the unreliable witness. Experimental evidence has yet to consistently support this proposal. The issue may well reside in the quality of the testimony, a question not yet fully researched.

The police can modify line-up procedure to decrease the danger of choosing innocent suspects without impairing culprit identification. The most prominent has been the sequential line-up. In the simultaneous line-up, the witness views all the line-up members at the same time. In the sequential line-up, witnesses view them one at a time, deciding after each one whether he or she is the culprit.

In the simultaneous line-up, witnesses who decide to choose when the culprit is absent do so by choosing the person who seems most similar to the culprit. That person is too often the innocent suspect. In the sequential line-up, they cannot compare between line-up members, and therefore cannot follow this strategy. The result is far fewer mistaken identifications, with very little reduction in culprit identifications.

The major problem is that people believe that the sequential line-up, actually far safer, is actually more dangerous. Despite the fact that the first paper demonstrating its superiority was published in 1985, it is widely used today only in the Canadian province of Ontario. Also, if only one line-up member can be chosen, some culprits escape prosecution. The witness may choose a foil before he sees the culprit.

A recently tested modified sequential line-up reduces mistaken identifications still more. This line-up has so far up to about 40 members, which naturally decreases the chances of the innocent suspect being chosen. Perhaps the line-up could be larger without reducing the chances of choosing the culprit.

In addition, witnesses may choose more than one line-up member; this further protects innocent suspects, as witnesses do so often when the culprit is absent. Even when the suspect is chosen and prosecuted, the court will realize that the 'identification' is a lot weaker, and will demand far stronger additional evidence.

Finally, the culprit has less chance of escaping identification: if witnesses choose a foil first, they can still choose the culprit later. Witnesses in experiments sometimes first choose the culprit, and then a foil. The culprit will also get extra protection in court, but that is only fair. The witness failed to differentiate a foil from the defendant, and the court should be aware of this important fact. The method has yet to be tested in the courts, but is in the process of being introduced in Israel.

Such a large line-up cannot be live. Rather than use photographs, the method relies on videoclips of line-up members. Some experiments have shown that video is as good as live. Video technology thus has the potential of reversing the trend towards photo line-ups, as videoclips can be saved and used repeatedly (in common with photographs).

The improvements discussed so far are aimed mainly at reducing the rate of false identifications, rather than increasing the rate of correct ones. The police have no control over eyewitness conditions that cause failures to identify the culprit. The police

can modify the line-up method that causes the false identifications.

Yet there are still line-up modifications that may increase culprit identification. We have noted the use of video line-ups instead of photographs. The principle is to maintain the maximum similarity between eyewitness conditions and the line-up. Another obvious extension would be to add the line-up member's voice to the videoclip. One experiment added the voice to still photographs and found marked improvement. However, distinctive accents could then act like different hair color in the line-up and would have to be avoided.

Another extension would be to conduct the line-up at the scene of the crime. Aside from creating practical difficulties, the experimental evidence is not supportive. An easier alternative would be to have the witness imagine the scene. The experimental evidence is inconclusive.

The Mugshot Search

The problem with the mugshot search today is that it is like a search for a needle in a haystack. When the method was first introduced, cities were smaller and so were mugshot albums. Today, mugshot albums are so large that witnesses can no longer peruse all the photographs and stay awake.

The essential requirement, then, is to select from the whole album a far smaller subset. The danger, however, is in throwing out the baby with the bath water. If the culprit is not in the subset, the exercise becomes futile. The challenge is to choose the smallest possible subset that includes the culprit.

The common police practice today is to use witnesses' verbal descriptions of the culprit to create such a subset. Thus, the police categorize each photograph in terms of such features as age, hair color and body build. The witness then views only those photographs that fit the description of the culprit, indeed a much smaller subset of photographs.

The method is not without its problems. Most of the features are facial, and some research has shown that witnesses forget these quickly. Also, there is much room for error. People do not necessarily look their age. Concepts such as 'thin body build' have fuzzy boundaries. One person's thin body is another's medium build. Finally, witnesses err even on clearer concepts, such as eye color.

There is also mounting evidence that people do not remember a face by its parts, but rather by how it looks in its entirety. In addition, there is research evidence to show that even giving a verbal description of the culprit may interfere with the witness's ability to later recognize him or her.

An alternative method would be to categorize photographs on the basis of their overall similarity, which would be more in keeping with recognition processes. A number of computerized techniques are available that can accomplish this task. However, two issues remain. First of all, the test of these methods has been to find a person's photograph using a different photograph of the same person. The definition of similarity is the objective definition involving a comparison between the two photographs. The mugsearch, however, involves a subjective definition, a comparison between the memory of the witness and a photograph. Programs tested using the former definition may not be effective in accomplishing the latter task.

Second, an appropriate task must be found to connect the computer program and the witness. One suggestion has been to have the witness construct a composite, and have the computer find the photographs similar to it. However, we have noted that the composite is only a likeness of the person, not a photograph. It remains to be seen whether this will be good enough.

Another suggestion has been somehow to make direct use of similarity judgments of the witness. In one study the witness viewed the highest ranked 24 photographs at a time, choosing the most similar to the 'culprit'. Using the similarity network, the computer pushed up in the ranking those photographs that were similar to those chosen. In this fashion, the culprit was pushed up and viewed much faster. However, the study did not test actual identifications, and another study has indicated that viewing photographs similar to the culprit will interfere with the identification. We note again the fragility of eyewitness evidence.

The Composite

The composite, interestingly, has been researched far more than the mugshot search. The most common experimental test of composites has been to put the photograph of the person whose composite was constructed among a set of other photographs. People then attempted to choose the correct photograph from the set. Experiments using this test have reached the conclusion that composites are not very good likenesses of the person. Only about 1 in 8 have been correct guesses.

Clearly, composite construction is a difficult task. The witness must first recall and then communicate to the composite constructor the various facial features. However, people are adept at neither recalling nor describing faces. The task of simply recognizing a face, required for the line-up and the mugshot search, is clearly much simpler.

One hypothesis explaining the poor performance of the composite was the inadequacy of the early, manual composite kits. The numbers of exemplars of each feature were limited, as were the means of modifying them. However, the few tests made with computerized composite programs, which have corrected these limitations, have not demonstrated better success rates.

In addition, all composite methods to date select parts of faces. As we have noted, this does not seem to be the way that people remember and recognize faces. An alternative approach would have witnesses choose photographs that are similar to the culprit. Superimposing these photographs could then create an average face.

While promising, such a technique returns us to a variant of the mugshot search to find those photographs that are similar to the culprit. The quality of the composite depends on the similarity to the culprit of the photographs found. Once we solve the mugshot search problem, adding the superimposition composite method would be simple. However, as we have noted, the mugshot search problem has yet to be solved.

Another potential avenue has been the use of caricatures, a proposal quite at variance with the superimposition idea. Superimposition involves averaging the features of a number of photographs, which of necessity results in a more average face. The caricature, on the other hand, emphasizes what is different, and creates a more distinctive face. So far, the experimental evidence evaluating this technique has been inconclusive; however, this may have resulted from lumping together different methods not equally effective.

None the less, police departments would consider a success rate of 12.5% (1 in 8) very good indeed. The composite is, after all, the last resort identification technique. The experimental test does not actually predict rate of success in the field. There are three major differences. First of all, the witness who constructs the composite in the experiment has calmly viewed the 'culprit's' photograph. We have noted, however, that real world witnesses view culprits under far more demanding conditions.

On the other hand, in the experiment strangers to the 'culprit' are asked to recognize him or her from a set of photographs. In the real world, the police count on finding someone who knows the culprit to identify him from the composite. The experimental task is more difficult. Finally, in the real world the composite must reach the eyes of someone who knows the culprit. This may be the most difficult requirement of all.

The rate of success in the real world is only 2%. That is, only 2% of composites actually aid in some way in apprehending a suspect who later is convicted in court. In most cases in which composites are constructed, either the culprit is apprehended by other means or no suspect is apprehended at all. Furthermore, when the experimental test was used, real-life composites tended to be poorer lookalikes than those constructed in experiments. That was true also for those 2% of composites that actually aided the investigation. This suggests that real-life eyewitness conditions are indeed more difficult. Finally, composites have little likelihood of reaching people familiar with the culprit: usually only police officers see them.

In Conclusion

There is much room for improvement in today's forensic tools for facial identification. Fortunately, new technology is now available for the line-up, the most crucial tool. Even in that area, however, the latest developments were only reported in 1998, and practice lags far behind advances reported more than 10 years earlier.

The fragility of the evidence, human eyewitness memory, is at the root of the difficulty. Time and again the methods devised to uncover the evidence have tended to destroy its value.

In addition, the forensic expertise of practitioners lags tremendously behind academic knowledge. Facial identification has yet to be acknowledged widely as a branch of forensic science. Very few practitioners in this field have professional academic degrees. Police continue to conduct line-ups that are either clearly unfair or far more dangerous than necessary. The challenge in the area of eyewitness identification today, then, is twofold: to achieve additional scientific breakthroughs; and to get these advances into the field.

(Figure 3 is a composite of the bottom right-hand member of the photographic line-up in Fig. 1.)

See also: **Hair**: Comparison: Other. **Facial Identification**: Photo Image Identification; Computerized Facial Reconstruction; Skull-photo Superimposition; Facial Tissue Thickness in Facial Reconstruction.

Further Reading

Brandon R and Davies C (1973) *Wrongful Imprisonment*. London: Allen and Unwin.

Brooks N (1983) *Pretrial Eyewitness Identification Procedures*. Ottawa: Law Reform Commission of Canada.

Ceci SJ, Ross DF and Togla MP (eds) (1987) *Children's Eyewitness Memory*. New York: Springer-Verlag.

Cutler BL and Penrod SD (1995) *Mistaken Identification:*

The Eyewitness, Psychology and the Law. New York: Cambridge University Press.
Davies G, Ellis H and Shepherd J (1981) *Perceiving and Remembering Faces*. London: Academic Press.
Devlin Lord (1976) *Report to the Secretary of State for the Home Department on Evidence of Identification in Criminal Cases*. London: Her Majesty's Stationary Office.
Ellis HD, Jeeves MA, Newcombe F and Young AW (eds) (1986) *Aspects of Face Processing*. Dordrecht, Netherlands: Martinus Nijhoff.
Gruneberg MM, Morris PE and Sykes RN (eds) (1978) *Practical Aspects of Memory: Current Research and Issues*. London: Academic Press.
Gruneberg MM, Morris PE and Sykes RN (eds) (1988) *Practical Aspects of Memory: Current Research and Issues*. Chichester: John Wiley.
Lloyd-Bostock SMA and Clifford BR (eds) (1983) *Evaluating Witness Evidence*. New York: Wiley.
Raskin DC (eds) (1989) *Psychological Methods in Criminal Investigation and Evidence*. New York: Springer.
Ross DF, Read JD and Toglia MP (eds) (1994) *Adult Eyewitness Testimony: Current Trends and Developments*. New York: Cambridge University Press.
Shepherd JW, Ellis HD and Davies GM (1982) *Identification Evidence: A Psychological Evaluation*. Aberdeen: Aberdeen University Press.
Wells GL and Loftus EF (eds) (1984) *Eyewitness Testimony: Psychological Perspectives*. London: Cambridge University Press.

Photo Image Identification

M Y İşcan, Istanbul Ünivesitesi, Istanbul, Turkey
S R Loth, University of Pretoria, Pretoria, South Africa

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0764

Introduction

A triple murder was committed by two men. The criminals entered through the back door of an appartment and surprised the owner and his friends. After a protracted search of the apartment, they beat, interrogated and eventually killed the victims in succession with large guns. Police arrived at the crime scene and noticed that there was a video surveillance camera similar to those used in a bank. To their surprise the entire event, lasting nearly half an hour from the time the criminals entered the house, was recorded on videotape. Using photo images that were isolated from the tape, the police arrested two suspects. Both suspects challenged the arrest and one of their defense attorneys called in an expert to analyze the tape.

In another incident, a man entered a bank and demanded money. The teller was advised to cooperate and she handed over all the money in her drawer. This event, too, was recorded by a video camera. Police apprehended and arrested a suspect. The terrified teller was shown the suspect and identified him as the man who robbed the bank. All of the other witnesses on the scene disagreed and positively declared that the suspect was not the robber. The suspect, a homeless man, denied committing the crime and insisted he was with a dying friend at the time of the robbery, but the friend, also homeless, could not be located. In this case, too, the defense attorney called in experts to analyze the videotape. In both cases, the prosecution also brought in their forensic experts to 'prove' that the suspects matched the criminals in the surveillance tapes.

In the first case, discrepancies demonstrated by video superimposition contributed to a hung jury; in other words, the jury could not agree on a verdict. In the second, a pretrial meeting during which the forensic anthropologists for the defense discussed their findings with the prosecutor led to charges being dropped against the suspect 1 day before the trial was scheduled to begin.

Events such as these are becoming frighteningly common. It is also becoming more common that these crimes are caught on videotape in a commercial establishment or by an ordinary citizen. As more and more surveillance cameras are installed in both public and private venues, the need to develop reliable methods of photographic comparison becomes acute. Methods must not only contain standards of assessment but also be adaptable, because each case presents a unique set of challenges. For one thing, criminals often attempt to disguise themselves with hats, sunglasses and the ever-popular ski mask. Yet the biggest problem is the highly variable nature and quality of surveillance photographs or tapes. In general the quality is poor for many reasons, including film speed, camera distance and angle, lens length and type, and lighting. Another detrimental factor is the reuse of tapes. While this common practice makes economic sense under normal circumstances, the cost can be enormous when it interferes with law enforcement's ability to identify and convict a criminal. Dark, shadowy, grainy tapes can make identification extremely difficult and can lead to criminals being

freed on the basis of 'reasonable doubt' if no other evidence exists. A frequent outcome is that the photographic evidence is too poor to be a definitive 'witness', and thus this potentially invaluable direct evidence cannot meet the necessary burden of proof. The case must rest on other, often less powerful, circumstantial evidence.

Ironically, it is often the poor quality of the video output that leads prosecutors and defense attorneys to consult a forensic anthropologist. This must be a scientist who is qualified to express expert opinions on many aspects of human biology and can quantitatively and qualitatively analyze and compare images. These can include both skulls and faces in photographs and video footage. Although it seems deceptively simple, this kind of comparison is one of the most difficult to make.

Until fairly recently, cases like this were relatively rare and no standard methodological procedures were in place: investigators simply developed or 'reinvented' techniques of their own to fit the situation. The purpose of this article is therefore to discuss technical and methodological problems, new research and development, and current methodology that has been used with repeatable success in comparing facial images. These approaches include facial morphology, photoanthropometry and video superimposition. Research, experience, cases studies and trial work have demonstrated that these approaches can be applied to reach a conclusion, be it positive, negative or indeterminate. It has also brought to light the weaknesses inherent in this kind of analysis and sparked further research in the field.

Technical Considerations

When consulting on any case of this nature, it is vital to ascertain the source of the photographic evidence. It is always preferable to use the originals because even a good copy will lose some detail. Law enforcement agencies often make enlargements or enhance the original to bring out detail and facilitate finding a suspect at the beginning of the investigation. However, during the process of image analysis, an enhanced photograph should not be used for comparison or superimposition because the processing can result in distortion. This was evident in a case where the spotlights in the original were round, while the same fixtures appeared oval in the enhanced still. Obviously, the facial contours and features would be similarly affected.

Many aspects of the equipment itself should be considered, such as the distance and angle of the surveillance camera with that of the subject being photographed. The type of lens in the camera can also affect perspective, as with a very wide-angle or fish-eye lens that is sometimes used to cover a large area. In general, an image may be elongated if the subject–camera distance is short. Greater distance might make a face look rounder. Another aspect one has to consider is the angle of the images. In a full frontal view, angulation would affect the height of the face and features like the nose. A face positioned near the standard Frankfort horizontal plane would be longer than one tilted upward or downward. In profile, the length of the nose or face itself is not affected by head tilt. While facial width dimensions would generally not be altered by moderate up or down tilt, they are sensitive to even slight lateral deviations.

Care must also be taken in the choice of photographs used for comparisons that have been provided by the family of the individual in question. If they were done professionally, there is always the possibility that the prints were retouched. This, of course, would make them inappropriate for a comparison because the purpose of retouching is to improve appearance by eliminating defects or softening imperfect features. In addition, even makeup can interfere with the process by obscuring scars or creating a younger look. Ironically, it is just these characteristics that can be the all important factors of individualization which are necessary for positive identification.

Morphologic Approach

The human face is a reflection of the individual uniqueness of a person. There can be hints to many aspects of the persona – everything from personality and temperament to overall health and levels of stress. Biologically, the facial phenotype is a product of genetics and environment that reflects features of populations in specific regions. Local populations are products of long-term inbreeding within the community as well as with nearby neighbors. To exemplify this, the so-called central European Alpines are described as basic round-headed Caucasoid brunettes with medium width noses. Their primary sorting criteria include dark to medium brown hair and eye color, globular heads with high foreheads, a mesorrhine nose with a slightly concave or straight profile and fleshy, 'blobby' and often elevated tip. They also have olive (brunette) white skin color, abundant hair and beard, round or square face with prominent gonial angles. Although such persons may now be seen on every continent, they will be much more likely to be indigenous to Europe and their frequency is expected to be higher in that region.

During the nineteenth and twentieth centuries, scientists developed remarkable research techniques for analyzing human facial variation. Practically

every detail of the face has been systematically divided into gradations and categories based on relative size, shape, presence or absence. In addition, some of the more complex facial features were simplified to allow more objective evaluations. For example, the facial profile can be assorted by projection, e.g. midface jutting with nose protruding, as opposed to chin or forehead jutting, and midface concave. Full face outlines can also have many variations, such as elliptical, round, oval, square or double concave around the zygomatic region, among others. **Table 1** shows some of the commonly observed facial features and their respective scales of observation. While the table was originally developed to observe a living person, some of the same observations can be made on prints captured from video or movie footage, and photographs. The evaluation of the face and its classification is not a simple matter and requires considerable experience of anatomy and understanding of human variation. The facial features listed in the table also provide an opportunity for the investigator to systematically examine the entire face in detail.

The human face is a dynamic structure and can transmit a wide range of expressions, from very minute to greatly exaggerated. Even the most subtle changes in expression may create a different perception in others. Emotions such as happiness, curiosity, concern, anger, fear, worry or surprise are instantly recorded and can just as quickly disappear. Physical and more permanent alterations result from aging, disease, weight gain/loss, graying and loss of hair or beard, and exposure to the sun. Studies have reported that smoking can accelerate wrinkling. Over a lifetime, all these external and internal factors leave their impressions and create a nearly infinite range of individual variation, both within and between people. These become important factors, especially when comparing images taken over time.

During growth, facial features change noticeably. In the face they include changes that result from the enlargement of the head, nose, jaws and eruption of teeth. **Figure 1** shows physiognomic changes from midchildhood (6.5 years) through late adolescence (18 years). Most obvious is the elongation of the lower half of the face and nose. Yet the pointed chin shape remains consistent. This stems from the fact that the bony foundation of this region must be established early in life to support the development and eruption of the deciduous and permanent dentition. The incisors (and permanent first molars) are the first adult skeletal components to form in the body, and most of the permanent anterior teeth are either well in progress or completely erupted by about the age of 7 years. Not only does this affect the overall height of the face, but research has demonstrated that the mandibular symphysis takes adult shape at about that time, and from then on merely increases in size. Eyebrow form and density shows little variation throughout this 12 year period. It is to be expected that males would develop bushier brows during adolescence, as well as a beard. There were no obvious new folds and creases in the face that can be associated with growth. The forehead height and the location and shape of the hairline also appeared unchanged. Ear forms are clearly visible in several pictures and the ear lobe is attached. The picture at age 17, in the middle of the bottom row, differs from others as it shows some retouching and smoothing of creases. Hair form also remained straight during this time, even though the retouched image shows a curly hair style. In this case the change resulted from a perm, but it is not unusual for hair form to change during adolescence (e.g. a straight-haired child can become curly, or vice versa). In any event, hair style is too easy to alter to be used as a primary criterion for photographic comparison. Actual form, texture and color can only be determined by microscopic analysis of the hair itself.

It is an indisputable fact that skin develops wrinkles with age. However, extreme variability in onset and progression makes it impossible to quantify the relationship of wrinkling patterns to age. It has always been a popular assumption that the chin and the eyes are the first places to look for the presence of wrinkling. **Figure 2** shows wrinkle and fold patterns in a face. As noted above, causative factors are numerous. The most obvious changes are the formation of new lines and the deepening of furrows. Subtle changes include asymmetry of the eyelids, which occurs when one loses its elasticity earlier than the other. This loss of tone often gives the impression of an epicanthic fold in a Caucasoid. Another factor is that cartilage can grow throughout life, leading to lengthening of the nose and enlargement of the ears. This can be particularly troublesome when attempting to compare a photograph of a man in his twenties with his own picture some 40 years later.

A primary focus of facial identification research is to isolate features that can be considered as factors of individualization. Such features may not be very easy to detect, especially in a relatively closed population where interbreeding is the norm. In that situation, many individuals share similar features. It is also true that, while populations share some obvious attributes, everyone has features that make him or her distinctive. It is these features that we rely upon in our everyday lives to recognize each other, even though we may not be consciously aware of the process. Even children will embarrassingly blurt out that someone has a long nose, big mouth or ears that 'stick out'. It is

Table 1 Morphological characteristics of the head and face observable on photographs and living persons

- Facial forms
  - Elliptical
  - Round
  - Oval
  - Pentagonal
  - Rhomboid
  - Square
  - Trapezoid
  - Wedge-shaped
  - Double concave
  - Asymmetrical
- Facial profiles
  - Jutting
  - Forward curving
  - Vertical
  - Concave
  - Lower jutting
  - Upper jutting
- Forehead height
  - Low
  - Medium
  - High
- Forehead width
  - Small
  - Medium
  - Broad
- Skin color
  - Pale
  - Brunette
  - Brown
  - Chocolate
  - Black
- Vascularity
  - Slight
  - Average
  - Extreme
- Freckles
  - None
  - Few
  - Moderate
  - Extreme
- Moles
  - None
  - Few
  - Moderate
  - Extreme
- Hair form
  - Straight
  - Low waves
  - Deep waves
  - Curly
  - Frizzy
  - Woolly
- Texture
  - Fine
  - Moderate
  - Coarse
  - Wiry
- Baldness
  - Absent
  - Slight
  - Advanced
  - Complete
- Beard quantity
  - Very little
  - Small
  - Average
  - Hairy
- Hair color: Head and beard
  - Black
  - Brown
  - Red bright
  - Golden
  - Red
  - Grey
  - White
- Red pigment
  - Absent
  - Present
- Iris color
  - Black
  - Brown
  - Green-brown
  - Blue-brown
  - Green
  - Grey
  - Blue
  - Other
- Eyefolds
  - Absent
  - Internal
    - Slight
    - Average
    - Developed
  - Median
    - Slight
    - Average
    - Developed
  - External
    - Slight
    - Average
    - Developed
- Palpebral slit
  - Down
  - Horizontal
  - Up slight
  - Up moderate
  - Up extreme
- Opening height
  - Small
  - Medium
  - Large
- Upper lid
  - Low
  - Medium
  - High
- Eyebrow thickness
  - Slight
  - Small
  - Average
  - Large
- Concurrency
  - Absent
  - Slight
  - Average
  - Continuous
- Eyebrow shape
  - Straight
  - Wavy
  - Arched
- Eyebrow density
  - Sparse
  - Thick
  - Bushy
- Nasion depression
  - Trace
  - Slight
  - Average
  - Deep
  - Very Deep
- Bony profile
  - Straight
  - Concave
  - Wavy
  - Convex
- Bridge height
  - Small
  - Medium
  - High
- Bridge breadth
  - Very small
  - Small
  - Medium
  - Large
- Tip thickness
  - Very small
  - Small
  - Average
  - Thick
- Tip shape
  - Pointed
  - Bilobed
  - Angular
  - Rounded
  - Blobby
  - Snub
- Septum tilt
  - Upward
  - Up slightly
  - Horizontal
  - Down slightly
  - Downward

Nostril form
- Slit
- Ellipse
- Intermediate
- Round

Nostril visibility
- Lateral
  - None
  - Slight
  - Medium
  - Visible
- Frontal
  - None
  - Slight
  - Medium
  - Visible

Nasal alae
- Compressed
- Slight
- Flaring
- Extended

Lip thickness
- Very thin
- Thin
- Average
- Thick

Eversion
- Slight
- Small
- Average
- Everted

Lip seam
- Absent
- Slight
- Average
- Present

Integument lips
- Thin
- Average
- Thick

Philtrum size
- Small
- Wide

Philtrum shape
- Flat
- Deep
- Sides parallel
- Sides divergent

Upper lip notch
- Absent
- Wavy
- V-Shape

Mouth Corner
- Straight
- Upturn
- Downturn

Alveolar prognathism
  - Absent
  - Slight
  - Medium
  - Pronounced

Malars
- Anterior projection
  - Absent
  - Slight
  - Medium
  - Pronounced
- Lateral projection
  - Compressed
  - Slight
  - Medium
  - Pronounced

Chin projection
- Negative
- Neutral
- Slight
- Average
- Pronounced

Chin type
- Median
- Triangle
- Bilateral

Chin from front
- Small and round
- Wide and round
- Pointed

Chin shape
- Dimple
- Cleft
- Double chin

Gonial eversion
- Compressed
- Slight
- Moderate
- Everted
- Very Everted

Ear size
- Small
- Medium
- Large

Ear projection
- Slight
- Medium
- Large

Helix
- Flat
- Slight roll
- Average
- Very rolled

Anti-helix
- Slight
- Medium
- Developed

Darwin's point
- Absent
- Present

Lobe
- None
- Soldered
- Attached
- Free
- Long and free

Other features

Birth marks:

Moles:

Wrinkles:

Asymmetry

Fatness:

Mustache:

Beard

Sideburns

Trauma:

Surgery:

Scars:

Glasses:

From İşcan (1993).

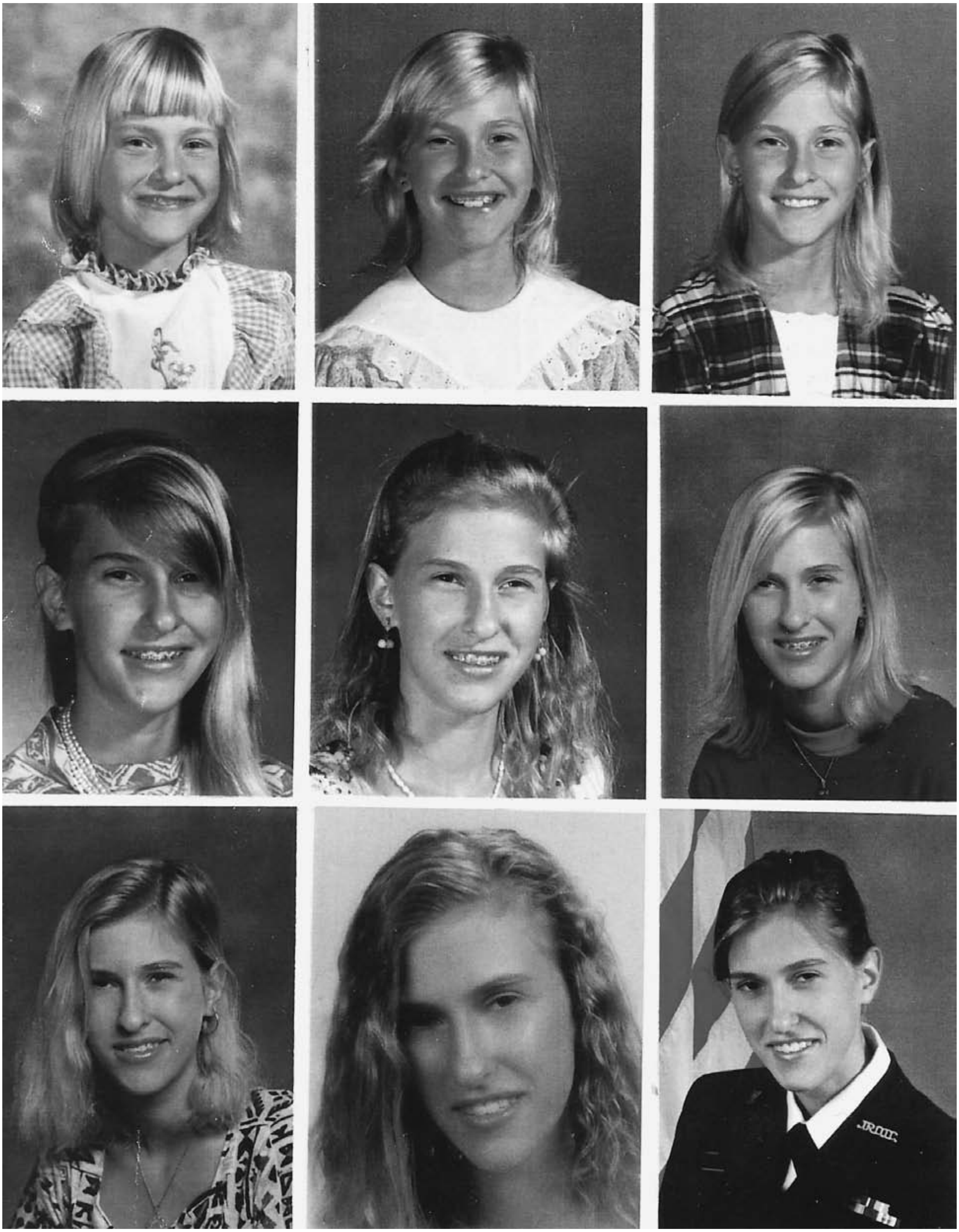


Figure 1 Pictures depicting physiognomic changes from middle childhood through late adolescence. Top row: ages are 6.5, 9.5 and 10.5 years; middle row: 12.5, 13.5 and 14.5 years; and bottom row: 15.5, 17 and 18 years.

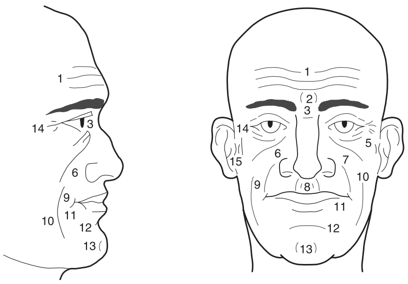


Figure 2 Wrinkle patterns: 1, horizontal forehead wrinkles; 2, vertical wrinkles of the glabella; 3, wrinkles of the nasal root; 4, eye fold below the orbit; 5, eye–cheek fold; 6, nose–cheek wrinkle; 7, nose–lip fold; 8, nose–lip fold; 9, cheek–chin wrinkle; 10, cheek–chin fold; 11, mouth corner fold; 12, lip–chin fold; 13, chin cleft; 14, temporal wrinkles; 15, ear wrinkles. Modified from Hammer (1978).

the scientist's job to isolate and identify even the most subtle differences in size, shape, asymmetry and proportion.

Every feature must be analyzed, for each gives clues at both the population and individual level. Hair color or form is important in race assessment. While Caucasoids can have nearly every color, texture and form, Mongoloids are characterized by straight, coarse dark hair and Negroids by dark, frizzy hair. Although hair can be easily colored, curled or straightened, the genetically programmed characteristics can be determined by microscopic analysis. Certain features point to populations within a given racial phenotype. Asian Indians, for example, are Caucasoid, but form a distinctly recognizable group. After sex has been determined, categorization of racial affinity is of primary importance. There must be sound scientific reasoning underlying the selection of features that are unique and unvarying enough to eliminate certain groups, and eventually individuals, as candidates. Attributes such as nose form, lip thickness and alveolar prognathism are among those structures known to show significant variation between races and are thus good structures with which to begin an analysis.

Admixture is increasingly common and can be both problematic and individualizing at the same time. Since the general public assigns racial affiliation based on the most dominant visible (phenotypic) characteristics, admixture may not always be apparent to casual observers and care must be taken not to eliminate candidates for identification. On the other hand, admixed features can be factors of individualization, such as an ostensibly black man with an unusually long narrow nose and little prognathism.

There are a number of considerations that should govern the choice of comparative sites. Often the selection is dictated by what is visible or clearest in the images. Examples of recommended sites include the eye, especially details like interpupillary distance, nasion or glabella, the tip of the nose, the base of the chin, ear shape and type of protrusion (upper, lower or both). It is advisable to choose structures that are most resistant to the ravages of time, cosmetic modification and weight changes. Easily modified features, especially those relating to hair (length and color, sideburns, beards and other facial hair), should be avoided. Even the hairline often recedes and changes shape during the adult lifespan.

In a recent attempt to classify human faces, 50 sets of photographs (full, profile and three-quarter views) of white male adults were analyzed. Each set was examined using 39 facial features selected and modified from Table 1. The researchers found that height and width dimensions, not defined by fixed points and requiring judgment by the observer, to be the most unreliable and unpredictable. Examples include forehead height and hair length. Features like face shape (e.g. oval or triangular) had higher rates of interobserver agreement, and pronounced ear projection was found to be one of the best discriminators.

Photoanthropometry

A second approach is based on the size and proportionality of various facial features. The technique traces its roots to traditional anthropometric methods. For the sake of clarity, it is referred as photoanthropometry. These quantitative approaches diverge when it comes to what is being assessed. Anthropometry is based on measurement of the living and deals with a three-dimensional structure. In the living, landmarks delineating measurements can be palpated and located precisely and repeatably with experience. When photographs are involved, the process becomes essentially two-dimensional. Photographic quality affecting the distinctness of lines and borders as well as the problems of perspective and depth of field are major factors that increase the difficulty of this analysis and lower the level of precision that can be expected. In addition, the forensic expert can only control the photographs he or she takes for comparison and, no matter how good those are, the limiting factor is the quality of the evidentiary photograph. Although the developing process can yield a brighter print or increase contrast, as noted above, care must be taken to avoid enhancements that cause distortions in size or shape.

A key element in the use of photoanthropometry is the formulation of indices based on proportions rather than absolute size. They assess the relationship of one structure (e.g. nose length) to another (e.g. total facial height). As in all analyses, the landmarks used for indices must be clearly visible and defined if they are not standard sites. An important consideration for

this approach is not to limit an analysis to preset or traditional landmarks. The quality and angulation of the image may dictate the use of unusual points that can be clearly defined and repeatably located on both images. Consistent replicability is essential, especially for courtroom presentations. This can be achieved by using a clear overlay to mark the reference points on each photograph without altering the evidence, obscuring features or interfering with other kinds of analyses.

To begin a metric photographic analysis the pictures should be copied and enlarged to approximate the actual size of the faces if possible. However, enlargements should only be made if they do not compromise the quality of the image. Once the landmarks have been chosen, they must be marked with a pen with a very fine tip on both photographs (but never on originals) to be compared. It is absolutely essential not to alter the originals. Suggested sites visible (in full face, profile or both) when the picture is oriented in the Frankfort horizontal plane are illustrated in **Figure 3** and are described as follows:

1. Trichion: midpoint of the hairline (full face and profile).
2. Metopion: most anterior point of the forehead (profile only).
3. Glabella: midpoint between the eyebrows on the median plane (profile only).
4. Nasion: deepest point of the nasal root (profile only).
5. Midnasal point: midpoint between the endocanthions (full face only).
6. Pronasale: most anterior point of the nose tip (profile only).
7. Subnasale: point where the nasal septum meets the philtrum (can be located in both full and profile if the nose tip is horizontal or elevated).
8. Superior labiale: midpoint of the vermilion seam of the upper lip (full face and profile).

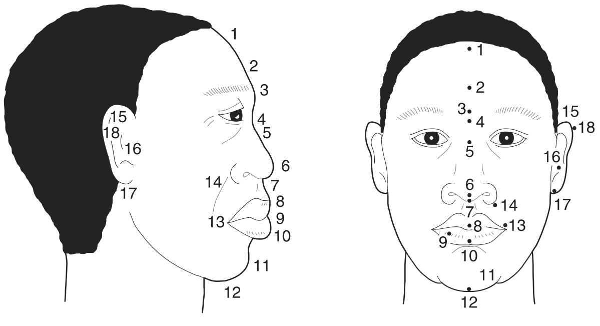


Figure 3 Anatomic landmarks of the face: 1, trichion; 2, metopion; 3, glabella; 4, nasion; 5, midnasal point; 6, pronasale; 7, subnasale; 8, superior labiale; 9, stomion; 10, inferior labiale; 11, pogonion; 12, gnathion; 13, cheilion; 14, alare; 15, superaurale; 16, tragion; 17, subaurale; 18, postaurale.

9. Stomion: midpoint of the occlusal line between the lips (full face and profile).
10. Inferior labiale: midpoint of the vermilion seam of the lower lip (full face and profile).
11. Pogonion: most anterior point of the chin (profile only).
12. Gnathion: Most inferior point of the chin (full face and profile).
13. Cheilion: corner of the mouth (full face and profile).
14. Alare: most lateral point of the nasal wings (full face and profile).
15. Superaurale: most superior point of the ear (full face and profile).
16. Tragion: most anterior point of the tragus (profile only).
17. Subaurale: most inferior point of the ear (full face and profile).
18. Postaurale: most posterior point of the ear (profile only).

Many measurements can be taken from these two views. If trichion (1) is clear, it can be the starting point for the majority of vertical measurements, e.g. 1–4 (1–5), 1–7 (or 1–6), 1–8, 1–9, 1–10, 1–12, 14–14 and 16–16. If trichion is not available or the hairline is receding, then 4 (or 5) can be the starting point. Facial width measurements can cover 1–16, 2–16, 4–16, 6–16, 7–16 and 15–17.

As mentioned earlier, the examiner is not limited to preset landmarks and others can be defined and used if they are better adapted to the images in question and are clearly visible in both photographs. Because of the variability in the type and quality of evidence each case presents, a standard set of dimensions cannot be established. All measurements must be taken with a precision caliper with vernier accurate to at least one decimal place.

Finally, since absolute size is not reliable without a scale, indices must be calculated from these measurements to insure that the values are comparable. Although standard-sized objects may be in the photograph and thus can be used as a scale, this rarely happens. In most cases it is impossible to determine the actual dimensions of the face and its features. Approximations are not recommended because the potential for error is too great. By relying on relative proportions, the index functions to eliminate the incomparability resulting from absolute size differences between images. An index is created as follows:

$$\frac{\text{Smaller dimension}}{\text{Larger dimension}} \times 100$$

For linear dimensions, it is best to use the maximum

dimension as a constant denominator (e.g. 4–7/1–12, 4–8/1–12, 4–8/1–12, etc.). Width or breadth can also be assessed in relation to height or other facial dimensions. Computer spreadsheets are recommended for this purpose because they can quickly and accurately generate the desired values.

This metric procedure is demonstrated using persons A and B in **Figure 4**. **Table 2** presents the values measured from the *original* drawings. Three measurements were taken for this exercise from both full and profile views. Since the actual sizes of these (fictitious) people are not known, three indices per facial view were calculated for this demonstration, but additional measurements can always be taken and indices computed to provide support for differences or similarities. The only restriction is that the points be clear on both pictures. It is obvious that person A has definite anthropometric differences from person B. Person B has a longer nose in relation to both facial height and breadth. He also has a wider face in relation to facial height. This was the case in both the frontal and profile images.

The marked photographs must be used for the investigation and submitted to the court as evidence. Copies should be made by reshooting prints after the photographs have been marked. This is the only way to duplicate exactly the positions of the marks. Photocopies should not be used for official purposes.

The aim of photoanthropometry is to compare metrically the proportional relationships of one photograph to another, rather than assess absolute visual differences and similarities as in morphological

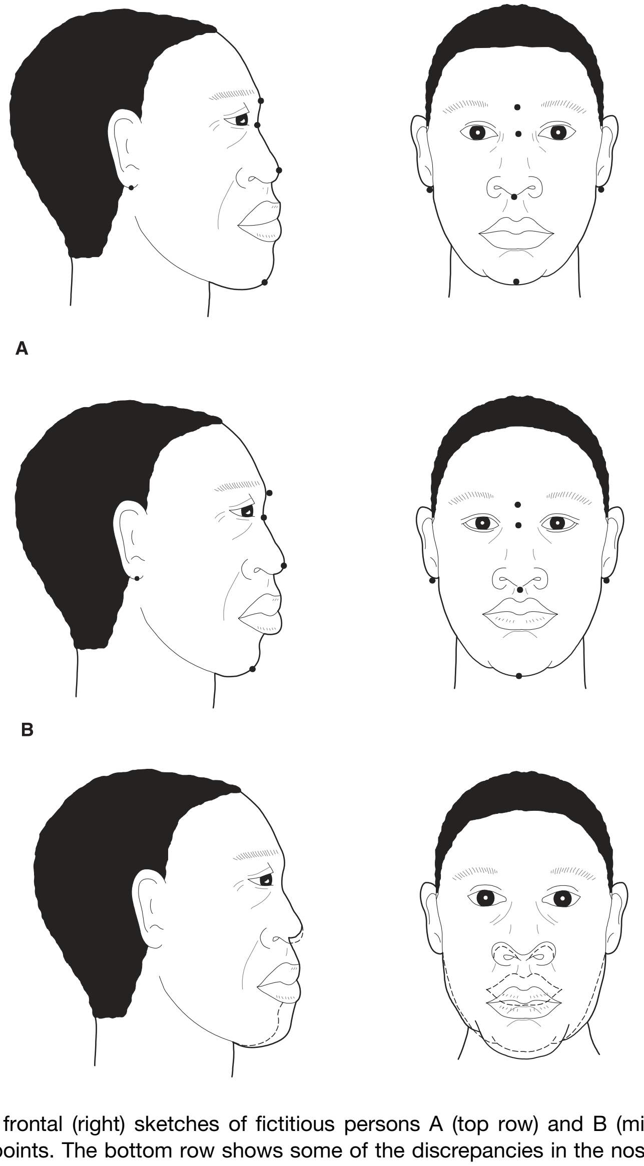


Figure 4 Profile (left) and frontal (right) sketches of fictitious persons A (top row) and B (middle row) for photoanthropometric comparison using marked points. The bottom row shows some of the discrepancies in the nose length, mouth position and lower facial border that would be seen when the faces are superimposed: —— person A; ---- person B.

Table 2 Comparison of indices calculated from the dimensions of frontal and profile views of person A (unknown) with person B (suspect)

Dimensions	*A*	*B*	*Indices*	*A*	*B*
Frontal view					
Glabella–Menton (face h)	39.4	36.1	Nose h/face h	34.3	38.8
Nasion–Subnasale (nose h)	13.5	14.0	Nose h/face w	36.4	38.1
Subaurale–Subaurale (face w)	37.1	36.7	Face w/face h	94.2	101.7
Profile view					
Glabella–Menton (profile face h)	40.0	37.6	Nose h/face h	28.0	32.2
Nasion–Pronasale (profile nose h)	11.2	12.1	Nose h/face w	34.4	38.5
Subaurale–Pronasale (profile face w)	32.6	31.4	Face w/face h	81.5	83.5

h, Height; w, width.

comparisons. While replicable quantification may reduce the subjectivity of photographic comparisons, one still cannot always reach a definitive conclusion regarding the matching of images. The most obvious problems are:

- The photographs are usually taken under different conditions.
- Proportional aberrations can be caused by differences in the lens angle and the camera-to-subject distance.
- Some photographs may have been retouched so that facial features may not match those of the unaltered photographs, even if it is the same individual.
- Weight and age differences can change the location of landmarks.
- Differences in facial expression may create differences in certain measurements (e.g. mouth size). This is especially true when a smiling visage is compared with a frowning one. The alterations of the facial musculature during laughter, for example, may widen the mouth, shorten nose-to-mouth distance and lengthen the face.

Further research in photographic analysis is needed to develop methods that can account for mensural variations, arising from differences in pose and angle, by experimenting with control subjects. Studies should follow exacting scientific protocols, including rigorous tests of empirical data. Experimentation along these lines has already begun.

Photographic Video Superimposition

Photographic superimposition is the method whereby two comparably enlarged photographs are superimposed using video cameras, a mixer and a monitor. Current video technology can eliminate the laborious work and expense of photographic enlargement and the difficulties of object orientation (e.g. a living person or a skull) to duplicate the pose of the picture. From a technical standpoint, the procedure can be readily accomplished with the right equipment. The challenging part is judging the quality of fit, identifying inconsistencies and knowing the range of acceptable variability.

The aim of this approach is to determine if facial features and dimensions match when two images are superimposed. Following proper placement of the cameras, photographic distances are adjusted to make the images the same size. At least three points must be aligned. These should include fixed points like the eyes, when visible, and one other clear recognizable part, like the nose, jaw line or contour of the head. A mixer then allows several types of visual comparison. In a 'vertical wipe' one face passes laterally across the other. The wipe can also be horizontal from top to bottom, and vice versa. A 'fade' makes one face disappear into another, with the second image eventually replacing the first. A spot analysis can be used to emphasize a selected region of the face. The spotlight can be focused on a particularly telling feature to call attention to differences or similarities. This should always be used to highlight an area of discrepancy, such as when the chin of one face extends beyond that of the other.

Fig. 4 (bottom row) illustrates the process of superimposition. The dotted lines (representing person B) highlight some of the discrepancies in the nose, mouth and lower face between person A (solid lines) and person B. It can be clearly seen that person A has a shorter nose, more protruding chin and narrower face.

As can be expected, there are a number of problems associated with superimposition, many of which also plague morphological and anthropometric comparisons. Of all types of superimposition, photo-to-photo is the least accurate. A problem unique to this comparison is the near impossibility of finding photographs that are taken under exactly the same conditions and with identical poses and expressions. The greatest accuracy can be achieved by superimposing a photograph on a living person or a skull, i.e.

three-dimensional objects. This allows the best possible alignment because the latter are not in a fixed position and can be easily adjusted to the orientation of the photograph. Another advantage is that the living person can imitate the facial expression in the photograph.

New applications of computer digitization have the potential to offer a more objective mode of superimposition. A computer can be programmed to quantify and evaluate the characteristics that differentiate one face from another. A program has been developed that digitizes projective symmetry in photographs and uses the resultant patterns to assess the comparability of two individuals. A composite image combining sections of each photograph is generated and if the components from both pictures fit perfectly, it is likely that the images are those of the same person. The developers report success even when a disguise is used. The latest concept in this field is experimentation with the application of a neural network system to compare two images. The aim is greater objectivity and the elimination of biases that may arise from individual observations. It also attempts to reduce the heavy reliance on symmetry that is the focus of digitization programs. This study even claims that age-related changes may be detectable with this technique. As with all new technology, rigorous independent testing is necessary before final judgments of effectiveness can be made.

For now, the best that can be said about technological advances concerns the very positive benefits derived from the use of video and computer systems that allow the expert to develop and demonstrate a series of visual effects and focus on morphological details. This in turn enables interested parties, like jurors, attorneys and judges, to see the entire procedure and visualize exactly how the experts reached their conclusions.

Expert Testimony

Being an expert witness is an awesome responsibility fraught with pitfalls that can have serious consequences. It is hard enough to carry out this job in areas with well-established criteria for judging evidence. Thus, the professionalism and experience of the witness becomes paramount in a developing field like this.

When an expert opinion is requested, the qualified scientist must demonstrate in detail how the analysis, superimposition for example, was carried out. The best approach is to allow plenty of time to show and explain how one photograph is superimposed over another. In most cases superimpositions do not match exactly, although a very slow fade or rapid wipe may give the illusion of a perfect match as one image is almost imperceptibly replaced by the other. Anything that can obfuscate inconsistencies must be avoided to maintain the integrity of the procedure. Following a complete initial wipe, the process of dissolving one image into the next should be repeated with pauses to focus on areas (highlighted by spot analysis if called for) where there are marked differences or excellent concordance. The expert should literally point out and trace these areas on the screen during pretrial deposition or testimony in court to insure that there is no misunderstanding.

An important aspect of being an expert witness is that one must pay scrupulous attention to any presentation by opposing experts. An expert must be prepared to advise counsel of errors or attempts to mislead. In some cases opponents focus only on what fits or try to distract by showing a nice overall agreement. For example, some witnesses draw attention to agreements in ear or nose length while ignoring obvious discrepancies in the chin and forehead. There are stringent requirements (addressed below) for declaring a positive match. An expert must not accept identification based solely on agreement in a few general features like facial roundness, projecting ears and receding hairline. One must also not lose sight of differences in expression between images such as an open mouth that is dissolving into a closed one or a smile that 'disappears'. Camera 'tricks' can either enhance or minimize variation between pictures. That is why we emphasize again that everything must be clearly described – what fits and what does not – no matter which side you represent. The same, of course, applies to metric and morphologic analyses. It can only hurt an expert's case and credibility to have the other side point out something that he or she failed to mention or attempted to obscure.

What conclusions are possible from photographic comparisons? Under optimal conditions an expert can make a positive identification. A definite declaration of positive identification is possible only if the following conditions are met: (1) all points match exactly with no discrepancy; (2) the images are clear and sharp at all points of comparison; and (3) there is a factor of individualization visible on both images. The third and absolutely essential feature can be a scar, gap between incisors, asymmetry, oddly shaped nose or head or other rare or unusual attribute that sets this individual apart. A match can also be ruled out. This conclusion can be reached when there are obvious discrepancies in the size or shape of attributes between the photographs. Similar-looking individuals can also be definitively differentiated by, for example, the presence of a scar in one image, but not the other. In these situations, the expert can confidently state

that the suspect is not the person photographed committing the crime. Lastly, most conclusions fall into the inconclusive or indeterminate category. This simply states that the expert cannot determine with 100% certainty that there is or is not a match. This situation can occur for a number of reasons, such as when there is a good match between the images but there is no factor of individualization. Although human variation is almost infinite, many people share certain features that are similar. Reaching definite conclusions is a real problem when someone has so-called 'average' features. This designation can also apply when one of the images is not very clear or sharp. In the authors' experience, the question of 'how possible?' often arises. In this situation, experts can use their discretion to indicate a stronger or weaker association, but there are no statistics on this type of probability, and no attempt should be made to guess at a 'ball park' figure. Most importantly, and as with any scientific pursuit, restraint must be exercised not to make any pronouncements that cannot be solidly supported by the data.

Discussion and Conclusions

The global rise in crime and the resultant video technology applied to fight it have created a growing need for anthropological expertise in image analysis. The three methods described in this article have been used in legal proceedings to compare one photograph with another (or videotape frame). They are: (1) morphological – detailed comparison of facial features following the form in Table 1; (2) photoanthropometry – quantitative analysis based on measurements of facial dimensions and the generation of indices based on them; and (3) photographic video superimposition – superimposition of one photograph or videotape frame over another. While it is rare to be able to make a positive pronouncement, two or more of these should be combined whenever possible to build a strongly supportive foundation for the final conclusion.

Of the comparative techniques presented here, the morphologic approach offers the best chance of obtaining a positive identification or ruling out a match, as is the case in skeletal analysis. It is the assessment of visible morphologic features that ferrets out factors of individualization, when present. Unfortunately for the investigator, many people simply do not have noticeable disfigurations or unusual facial elements. Most folks are 'average looking' and they inevitably fall into the 'indeterminate' category. Although the general public and experts alike clearly notice that people look different, they are often at a loss to describe exactly how two similar individuals differ. There is a vague sense that maybe one is a bit 'prettier' or more 'masculine'. What makes one individual different from all others is the highly varied complexity of many features and how they come together. It cannot be overstressed that a match must not be made on the basis of a general look, i.e. the common traits that define a racial phenotype, for example. Interpretations and conclusions must always be made with caution.

Photo-to-photo superimposition can provide an easy comparison of images and show discrepancies. It brings together two separate images and illustrates the proportional relationship of facial features in both photographs. Even if differences are distinct morphologically, or if there are clear factors of individualization, a superimposition may still show the location of these individualizing factors in the two images.

The methods discussed here do have the potential to conclude that there is a positive match, or rule one out decisively. However, it cannot be overemphasized that, no matter how good our methodology becomes, the process of comparison and superimposition is seriously compromised by the often extremely poor quality of surveillance photographs. In the absence of a skull or live subject that can be posed, the less than ideal condition of pictures representing missing persons further compounds the difficulties of analysis. This makes it difficult, if not impossible, to precisely place landmarks and borders for comparison. Another obstacle is the difficulty of determining absolute size. The probability of error is too great when one must extrapolate true dimensions in a photo with no scale. An invaluable improvement would be the installation of a measuring scale in the camera that would appear on each frame of the film. With the size factor accounted for, it would be possible to rule out many suspects with a few, simple, raw metric comparisons.

Finally, at the present level of development of this emerging specialty, conclusions in most cases will remain possible or indeterminate, at best. Therefore, photographic evidence often cannot be the decisive factor upon which to convict or exonerate a suspect. Technical and methodological difficulties must both be addressed for significant progress to be made in this type of identification. The expert witness must be thoughtful and responsible in making conclusions. Stretching the data to the point of conjecture must be avoided, and murky or doubtful areas presented as such. Credibility and cases can be lost unless experts fully comprehend and are able to communicate both the potential and limitations inherent in photographic identification.

See also: **Facial Identification**: Lineup, Mugshot Search and Composite; Computerized Facial Reconstruction;

Skull-photo Superimposition; Facial Tissue Thickness in Facial Reconstruction. **Identification/Individualization**: Overview and Meaning of ID.

Further Reading

Adams D (1991) Age changes in oral structures. *Dental Update* **18**(1):14–17.

Burke KE (1990) Facial wrinkles: prevention and nonsurgical correction. *Postgraduate Medicine* **88**(1):207–222.

Castelo-Branco C, Figueras F, Martinez dO and Vanrell J (1998) Facial wrinkling in postmenopausal women. Effects of smoking status and hormone replacement therapy. *Maturitas* **29**(1):75–86.

Catterick T (1992) Facial measurements as an aid to recognition. *Forensic Science International* **56**:23–27.

Doual JM, Doual-Bisser A, Crocquet M and Laude M (1991) Contribution à l'étude du viellissement facial: evolution des tissus mous. *Bulletin du Groupement International pour la Recherche Scientifique en Stomatologie et Odontologie* **34**(1):11–15.

Grady D and Ernster V (1992) Does cigarette smoking make you ugly and old? *American Journal of Epidemiology* **135**(8):839–842.

Hammer HJ (1978) Körperliche merkmale. In: Hunger H and Leopold D (eds) *Identifikation*, pp. 391–404. Leipzig: Barth.

Hooton EA (1946) *Up from the Ape*. New York: Macmillan.

İşcan MY (1988) Rise of forensic anthropology. *Yearbook of Physical Anthropology* **31**:203–230.

İşcan MY (1993) The introduction of techniques for photographic comparison: potential and problems. In: İşcan MY and Helmer RP (eds) *Forensic Analysis of the Skull: Craniofacial Analysis, Reconstruction, and Identification*, pp. 57–70. New York: Wiley-Liss.

Kadunce D, Burr R, Gress R, Kanner R, Lyon J and Zone J (1991) Cigarette smoking: risk factor for premature facial wrinkling. *Annals of Internal Medicine* **114**(10):840–844.

Koury ME and Epker BN (1991) The aged face: the facial manifestations of aging. *International Journal of Adult Orthodontics and Orthognathic Surgery* **62**:81–95.

Kritscher H and Szilvassy J (1991) *Zur Identifizierung Des Mozartschadels*. Annalen Des Naturhistorischen Museums in Wien, vol. 93, series A. Vienna: Selbstverlag Naturhistorisches Museum.

Krogman WM and İşcan MY (1986) *The Human Skeleton in Forensic Medicine*. Springfield, IL: Thomas.

Lan Y and Cai D (1993) Technical advances in skull-to-photo superimposition. In: İşcan MY and Helmer RP (eds) *Forensic Analysis of the Skull; Craniofacial Analysis, Reconstruction, and Identification*, pp. 119–129. New York: Wiley-Liss.

Linney A and Coombes AM (1998) Computer modelling of facial form. In: Clement JG and Ranson RL (eds) *Craniofacial Identification in Forensic Medicine*, pp. 187–199. London: Arnold.

Majumdar T and Sinha P (1989) Photographs of the human face and broken projective symmetry. *Journal of the Forensic Science Society* **29**(6):387–395.

Sinha P (1998) A symmetry perceiving adaptive neural network and facial image recognition. *Forensic Science International* **98**(1–2):67–89.

Vanezis P, Lu D, Cockburn J *et al.* (1996) Morphological classification of facial features in adult Caucasian males based on an assessment of photographs of 50 subjects. *Journal of Forensic Sciences* **41**(5):786–791.

Warren R, Gartstein V, Kligman AM, Montagna W, Allendorf RA and Ridder GM (1991) Age, sunlight, and facial skin: a histologic and quantitative study. *Journal of the American Academy of Dermatology* **25**(5 pt. 1):751–760.

Wogalter MS and Hosie JA (1991) Effects of cranial and facial hair on perceptions of age and person. *Journal of Social Psychology* **131**(4):589–591.

Yoshino M, Kubota S, Matsuda H *et al.* (1996) Face-to-face video superimposition using three dimensional physiognomic analysis. *Jpn J Sci Tech Iden* **1**(1):11–20.

Skull-photo Superimposition

M Yoshino, National Research Institute of Police Science, Tokyo, Japan

S Seta, St Marianna University School of Medicine, Kawasaki City, Japan

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0767

Introduction

Various skeletal parts provide valuable data on individual human characterization of unknown remains, particularly the skull. In criminal cases, the identification of skeletal remains is frequently carried out by skull–photo superimposition because a facial photograph can be easily obtained from the victim's family. The skull–photo superimposition technique is considered to be reconstruction, in that it attempts to supply a facial photograph for a discovered unknown skull. In order to identify someone through the superimposition technique, the examiner must investigate the outline and positional relationships of the skull to face parts, based on anatomical data after the skull to comparison photograph orientation is achieved.

From the forensic standpoint, the most instructive example of the application of this technique is provided by the two skulls from the Ruxton murder case

of 1935 in England. In this case, there were two female skulls, listed as skull no. 1 and skull no. 2. It was assumed that one was that of Mrs Ruxton and the other belonged to the housemaid Mary Rogerson. Comparisons using a superimposition technique in Mrs Ruxton's case demonstrated that skull no. 1 could not possibly be the skull of Mrs Ruxton, and that skull no. 2 might be the skull of Mrs Ruxton. In Miss Rogerson's case, the comparisons demonstrated that skull no. 2 could not possibly be the skull of Miss Rogerson, but that skull no. 1 might be the skull of Miss Rogerson. As a general rule, it can be stated that superimposition is of greater value in ruling out a match between the skull and facial photograph; however, researchers are led to believe that skull–photo superimposition is an efficient method for identifying the unknown skull because facial photographs are routinely used as a reference object of the victim. Of course, in some cases a facial photograph of another person may well be consistent with the skull in question in the outline and anatomical relationships between the skull and face. Therefore, forensic examiners must be well versed in the anatomy of the skull and face for effective utilization of the superimposition technique.

Anatomical Relationships between Skull and Face

When evaluating anatomical consistency between these parts, special attention should be paid to their outline, the facial tissue thickness at various anthropometric points, and the positional relationships between skull and face.

Thickness of facial soft tissue

The thickness of facial soft tissue is one of the important factors in reconstructing the face of an unknown skull. Several methods, such as puncturing, radiography, ultrasonic probing and computed tomography, are used for measuring soft-tissue thickness on the various anthropometric points of the head. Data on the thickness of facial soft tissue have been accumulated for each race: Caucasoid, Negroid and Mongoloid. **Tables 1** and **2** on pages 784 and 785 summarize this data. In the application of superimposition, examiners should first investigate the contours of both skull and face, and then must take into account soft-tissue thickness on various points and evaluate its individual variation (**Fig. 1**).

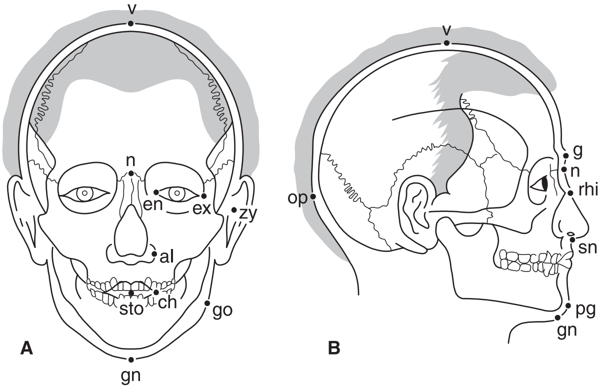


Figure 1 **(A)** Frontal and **(B)** lateral views of the head showing landmarks for the examination of consistency between a skull and a facial photograph by the superimposition method. al, Ala; ch, cheilion; en, entocanthion; ex, ectocanthion; g, glabella; gn, gnathion; go, gonion; n, nasion; op, opisthocranion; pg, pogonion; rhi, rhinion; sn, subnasale; sto, stomion; v, vertex; zy, zygion.

Positional relationships between skull and face

Relation of eyebrow to supraorbital margin The lower border of the eyebrow is generally located on the supraorbital margin. According to some researchers, the placement of the eyebrow is approximately 3–5 mm above the supraorbital margin.

Relation of eye to orbit In vertical placement, the eye slit (palpebral fissure), which is the line connecting the inner canthus (entocanthion) and the outer canthus (ectocanthion), falls approximately on the lower third of the orbital height (**Fig. 2**). The vertical level of the inner canthus corresponds approximately to the point of attachment of the medial palpebral ligament; this point is located about 10 mm below the dacryon, the junction of the lacrimomaxillary, frontomaxillary and frontolacrimal sutures. The position of the outer canthus corresponds to the point of attachment of the lateral palpebral ligament; this point is called the 'malar tubercle' and is located about 11 mm below the point at which the zygomaticofrontal suture crosses the orbital margin. The position of the outer canthus is slightly higher than that of the inner canthus (Fig. 2). In horizontal placement, the inner canthus is positioned more than 3 mm lateral to the medial orbital margin (approximately 5 mm), and

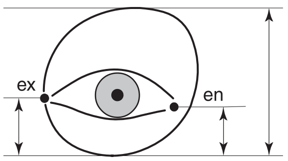


Figure 2 Relation of the eye to the orbit. The eye slit lies on the lower third of the orbital height. The entocanthion (en) lies more than 3 mm lateral to the medial orbital margin. The ectocanthion (ex) lies on the lateral orbital margin or inside it.

the outer canthus lies on or slightly within the lateral orbital margin (Fig. 2).

Relation of external nose to nasal aperture The height of the nasion-subnasale (external nose) corresponds to that of the nasion-subspinale on the skull, although the subspinale and subnasale do not lie at the same level. There is a difference in the two measurements favoring the external nose, which averages 1.4 mm in the adult Caucasoid and 1.6 mm in the adult Negroid. In the Mongoloid, the position of the rhinion corresponds to the upper third and the position of the nasospinale corresponds to the lower seventh of the nasion-subnasale height (**Fig. 3**). The measurement of the width of the external nose (ala to ala) always exceeds that of the nasal aperture. In general, alae lie 5 mm lateral to the lateral margin of the nasal aperture in the Caucasoid, 8 mm in the Negroid, and 6.5 mm in the Mongoloid (Fig. 3).

Relation of lips to teeth At rest, the oral slit appears to coincide with the line formed by the bite. The central point of the oral slit (stomion) falls 1–2 mm higher than the cutting margin of the upper central incisor. The oral angle (cheilion) generally lies on the junction of the canine and the first premolar. The average distance between the upper and lower teeth in repose is about 3 mm at both the first premolar and the first molar.

Relation of ear to porus acusticus externus The porus acusticus externus is generally located just behind the tragus. According to some researchers, the relation of the most lateral part of the cartilaginous portion to the bony portion of the ear tube is 5 mm above, 2.6 mm behind and 9.6 mm laterally.

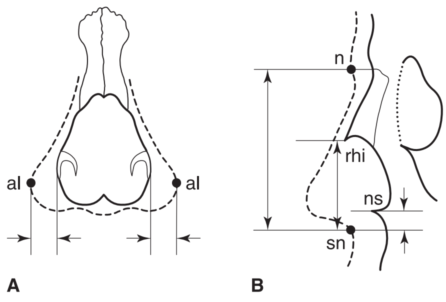


Figure 3 Relation of the external nose to the nasal aperture. **(A)** The average distance from the lateral margin of the nasal aperture to the ala (al) is 5 mm in the Caucasoid, 8 mm in the Negroid and 6.5 mm in the Japanese. **(B)** The position of the nasospinale (ns) corresponds to the lower seventh of the height of the external nose (n–sn).

Technical Development

From the aspect of technical development, skull–photo superimposition can be seen to have passed through three phases: the photographic, video and computer-assisted superimposition techniques. The photographic superimposition technique was developed in the mid-1930s. Since the latter half of the 1970s, the video superimposition technique has been widely used in the field of forensic anthropology and odontology. Computer-assisted superimposition was introduced in the latter half of the 1980s. In all techniques, the two most important factors are the enlargement of the facial photograph and the alignment of the skull to the photograph.

Photographic superimposition

When performing photographic superimposition, the examiner must always enlarge the comparison photographs to the size of the unknown skull and then position the skull in the same orientation as the facial photographs.

Enlargement of facial photograph Enlargement of facial photographs has been accomplished by employing measurable objects for the enlargement scale, facial proportion and distances, and anatomical landmarks on the skull and face. For instance, a magnification factor may be calculated from various measurements of the chair and the width of the collar in a photograph, and a life-sized facial photograph of the individual is reproduced. When the anterior teeth are seen in a facial photograph, the known sizes of anterior teeth in an unknown skull are used to decide the magnification factor for the facial photograph. In the use of anatomical landmarks, the frontal view facial photograph is enlarged from the original photograph based on bizygomatic skull width and soft-tissue thickness of the zygion. In the oblique or lateral view, the facial photograph enlargement is based on glabella–gnathion distance and the soft-tissue thickness of the gnathion.

Orientation of skull to facial photographs A craniophore used for anthropometry and a pivoting head tripod are most commonly used for positioning the skull. To date, a number of devices have been designed to position the skull so that it will adjust to the orientation of the facial photograph. The pulse motor-driven skull rest has been developed to solve the orientation problem. The skull rest can be moved by remote control levers in six directions: upwards and downwards, left and right, forwards and backwards, rotation around the axis, extension and flexion, and tilt to the left and right. The combination of

these movements adequately establishes the orientation of the skull to multipositional facial photographs. The skull rest is also used for both video and computer-assisted superimposition techniques.

Procedure for photographic superimposition 'Seeing through' techniques, in which orientation of skull to facial photograph is achieved by looking through a transparent face film, are optimal for photographic superimposition. The simplest method is to use the ground glass of the camera as a transparent plate. An original available photograph is taken and reproduced on quarter-size cut film. The negative thus obtained is placed under the ground glass of the camera, and the outline of the face is drawn using very fine pen and Indian ink. A 1 inch scale is placed on the forehead of the skull on a suitable skull rest, so that the skull may be arranged in the same perspective as the portrait. The skull is then photographed on quarter-size cut film. The negative is enlarged to natural size, so that the image of the 1 inch scale is exactly 1 inch, and then printed on bromide paper. A positive transparency is then taken on 24 × 30 cm X-ray film. Finally, the natural-size transparency of the skull and the life-size transparency of the portrait are superimposed and photographed on quarter-size cut film using transmitted light. An optical bench has been manufactured for the 'seeing through' method, consisting of a skull rest, a frame with criteria lines (rubber bands), a thin plastic plate and a camera. The skull is fixed at one side of the optical bench, and the camera at the other. The appropriate anthropometric points (e.g. nasion, gnathion, gonion, zygion, etc.) are marked on an enlarged facial photograph and then three guidelines are drawn, based on these points with India ink. Three criteria lines within the frame are adjusted to the guidelines on the facial photograph. The outline of the head and the guidelines are drawn on the thin plastic plate. When the criteria lines of the frame coincide with the guidelines on the thin plastic plate, the skull position is set, taking soft-tissue thickness into account. The enlarged facial photograph without the marker points and the guidelines is set on to the thin plastic plate, and then photographed. Finally, removing the frame and thin plastic plate, a double exposure of the positioned skull is taken.

In the 1980s, special apparatus for superimposition were devised. An example of such an apparatus reproduces the skull on the ground glass in half the original skull size via four plates of reflective mirror and a projection lens (**Fig. 4**). The comparison facial photograph is also reproduced at half the original face size according to anatomical landmarks. The image of the skull on the ground glass is then photo-

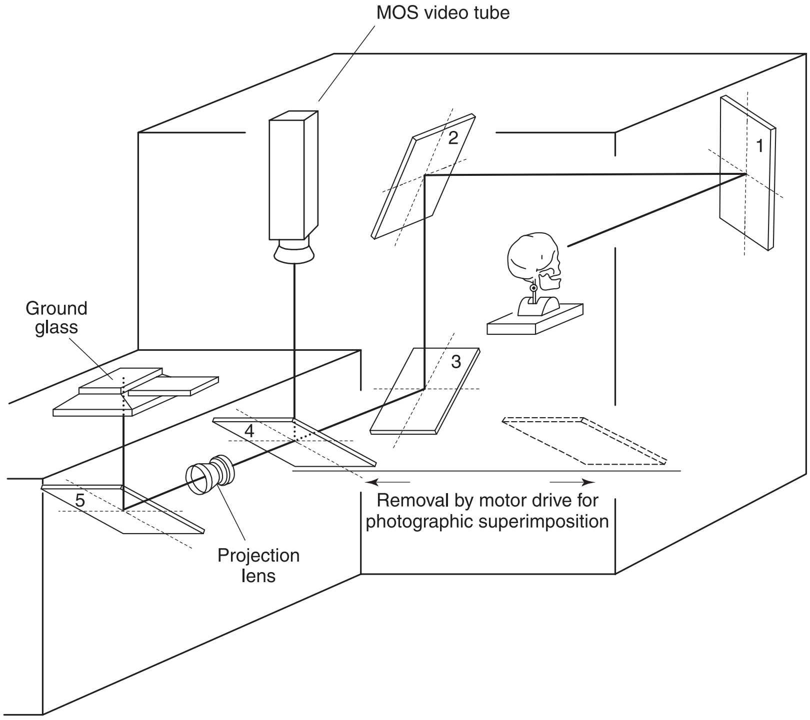


Figure 4 Illustration of the optical pathway. The optical pathway is produced by the combination of four reflection mirrors. For photographic superimposition, the fourth mirror is designed to slide out of the optical pathway in order to focus the skull image on the ground glass via a projection lens and mirror (5); for video superimposition mirrors 1–4 are used and the skull image is taken with a video camera. Reproduced from Seta and Yoshino (1993).

graphed using a cabinet-size film. The skull and face films are finally superimposed on the light box (**Fig. 5**).

Video superimposition

It has been suggested that video superimposition provides an enormous advantage over the usual photographic superimposition, in that comparisons can be made rapidly and in greater detail than is possible with static photographs. In general, the video superimposition system consists of a skull-positioning rest, two video cameras (1 and 2), a video image mixing device, a TV monitor and a videotape recorder (**Fig. 6**). The skull is first filmed with video camera 1 and reproduced on the TV monitor through the video image mixing device. The facial photograph is then reproduced on the TV monitor using video camera 2. Some laboratories use two video cameras, electronic and mixer units and three monitors. The skull is taken with camera 1 and the image is reproduced on monitor 1. The facial photograph is then reproduced with camera 2 on monitor 2. The mixed picture and horizontal and vertical sections are punched up on monitor 3 by means of the electronic and mixer units. To make the comparison, the skull image is superimposed on the facial image, and then adjusted to match the orientation of the face by the motor-driven mechanism described above. The size of the skull image is easily adjusted to that of the facial image using the zoom mechanism of camera 1, taking soft-tissue thickness into account. This technique permits fade-out and fade-in of the image of either the skull or face on the monitor for overall assessment of how well the two images match (**Fig. 7**). It also allows sectioning of the various images of the skull and facial images on the monitor (Fig. 7). The fade-out and wipe images, especially the latter, facilitate the comparison of positional relationships between the skull and face. However, in the video superimposition, the thickness of soft tissue at the anthropometric points and the distance between the anatomical landmarks cannot be directly evaluated on the monitor. The superimposed image on the monitor can be photographed directly with the camera or permanently preserved on videotape.

Computer-assisted superimposition

As computer technology advances, computer-assisted superimposition has become a popular method of identifying the unknown skull. Computer-assisted superimposition is divided roughly into two categories

Figure 5 **(A)** Frontal and **(B)** oblique photographic superimposition images, showing a good match between skull and facial photographs. Reproduced from Yoshino *et al*. (1995).

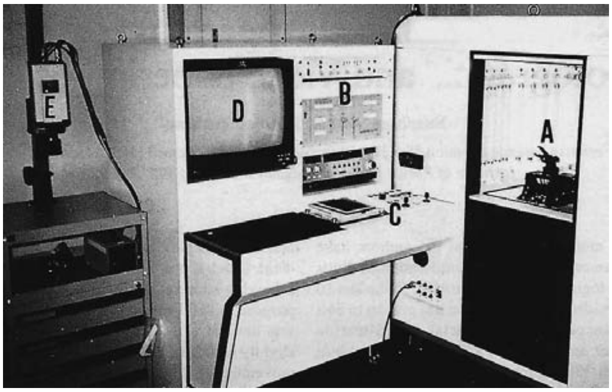


Figure 6 Video superimposition system. A, Skull-positioning box; B, electronic and mixing devices; C, control panel with joystick levers and momentary paddle switch; D, TV monitor; E, photo-stand for taking the facial photograph. Reproduced from Seta and Yoshino (1993).

from the point of view of identification strategy. The first method is to digitize the skull and facial photograph using a video computer with appropriate software and then to compare the two images morphologically by image processing. Establishing a scale for the digitized skull image is performed by converting the actual measurement between the landmarks (e.g. zygion–zygion, nasion–gnathion, etc.) into the number of pixels on the monitor. For the assessment of anatomical consistency between the digitized skull and face, the distance between the anatomical landmarks and the thickness of soft tissue at the anthropometric points are measured semiautomatically (**Fig. 8**). The software allows for fade-out

Figure 7 Video superimposition showing **(A)** total mixing and **(B)** vertical wipe images of the skull and facial photograph. Reproduced from Yoshino *et al.* (1997).

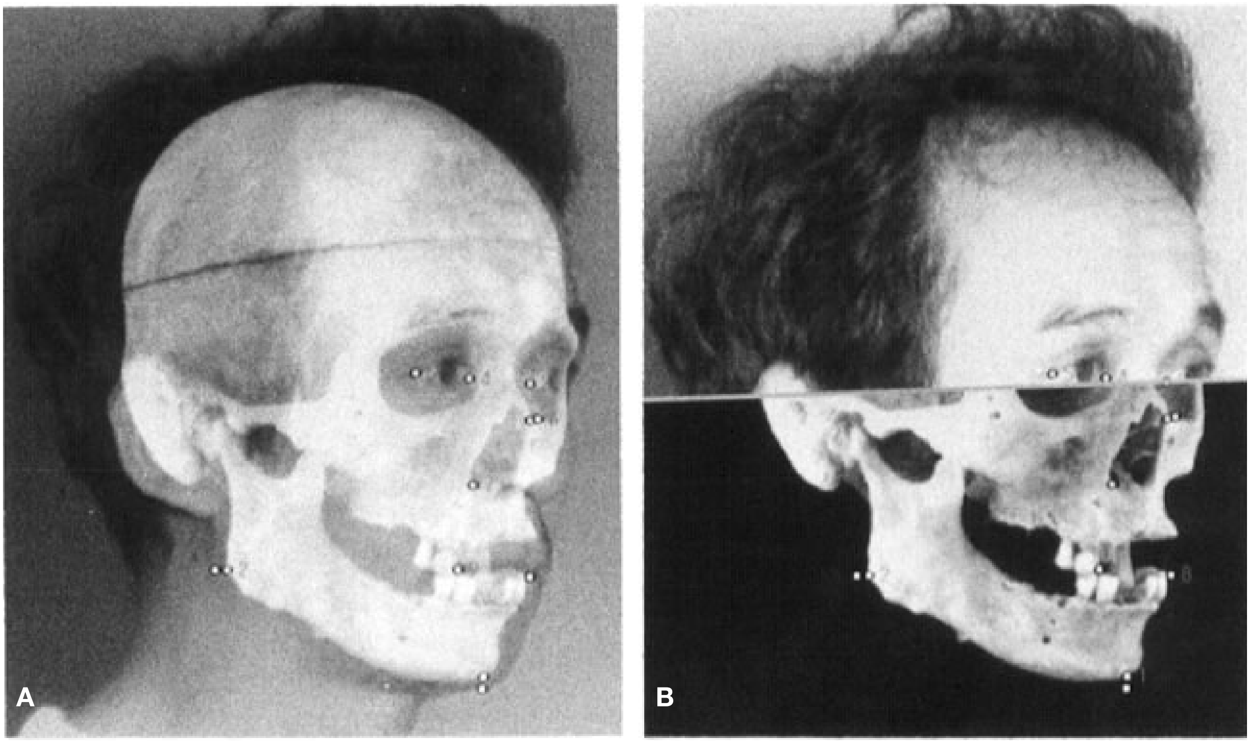


Figure 8 **(A)** Total mixing and **(B)** horizontal wipe images showing comparison between the digitized skull and face. The thickness of soft tissue of the anthropometric points and the distance between the anatomical landmarks are measured by means of pair-dots using the mouse. Reproduced from Yoshino *et al*. (1997).

and fade-in of the image of either skull or face, or wiping either the skull or face in the vertical or horizontal plane for evaluating anatomical relationships between them (Fig. 8). Some researchers have developed a methodology for comparing the skull and facial image using the near optimal fit between the three-dimensional skull surface mesh and the two-dimensional digitized facial photograph.

The second method is to evaluate the fit between the skull and facial image by morphometric examination. For instance, polynomial functions and Fourier harmonic analysis are applied to assess the fit between the outline of the skull and the face (**Figs 9** and **10**). If quantitative data based on morphometric analyses are obtained in addition to the anatomical and anthropometric evaluation, the judgment of skull identification will be more objective and reliable.

A new approach for judging the match between the skull and a facial photograph examines mathematically the coordinate values of the paired anatomical landmarks of the skull and face.

Evaluation of Consistency between Skull and Face

It is generally accepted that the unknown skull without the mandible cannot positively be identified as the presumed person, even if a good match is seen in the skull–photo superimposition image. The complete skull is required to obtain any degree of certainty.

The skull–photo superimposition should be able to prove whether or not the skull and facial photograph belong to the same person. In general, it is considered that the skull–photo superimposition is more valuable for exclusion purposes, because it may be definitely stated that the skull and facial photograph are *not* those of the same person. However, if they are well matched with each other, it can only be stated that the skull *could* be that of the photographed person. This is because of the possibility that this skull may be consistent with the facial photograph of another person, or another skull of the same size and contour may be consistent with the face of the photographed person. Experimental studies on the reliability of skull–photo superimposition suggest that a false match is almost always excluded when both frontal and lateral view photographs are compared with the skull. Unfortunately, it is often difficult to obtain a lateral view photograph in actual cases, and oblique view photographs are used to enhance the degree of consistency between the skull and face. Skull–photo superimposition must be considered highly corroborative, almost conclusive for personal identification when two or more photographs, clearly depicting the facial features from different angles, are used in the comparison. The outline from the forehead (trichion)

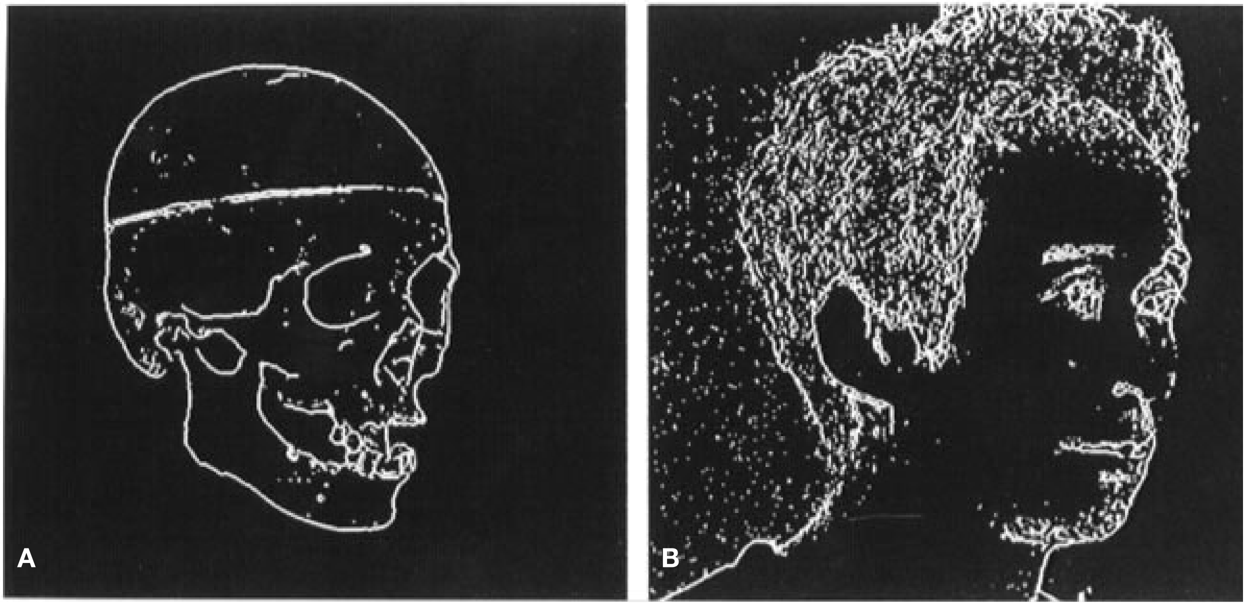


Figure 9 Thinning images extracted from **(A)** the digitized skull and **(B)** facial images. The forehead region of the thin line is selected for evaluating the fit between the contour of the skull and face. Reproduced from Yoshino *et al*. (1997).

to the gnathion in the lateral or oblique view is the preferred portion for identification.

When the incisors and canines can be seen in the facial photograph, they can be used as reliable data. The coincidence of dentition between the skull and facial photograph could lead to positive identification. Pathological findings or congenital abnormalities in the skull and facial photograph, in addition to points of consistency between the skull and photograph, certify that they belong to the same person.

See also: **Anthropology**: Morphological Age Estimation; Sex Determination; Determination of Racial Affinity. **Facial Identification**: Lineup, Mugshot Search and Composite; Photo Image Identification; Computerized Facial Reconstruction; Facial Tissue Thickness in Facial Reconstruction.

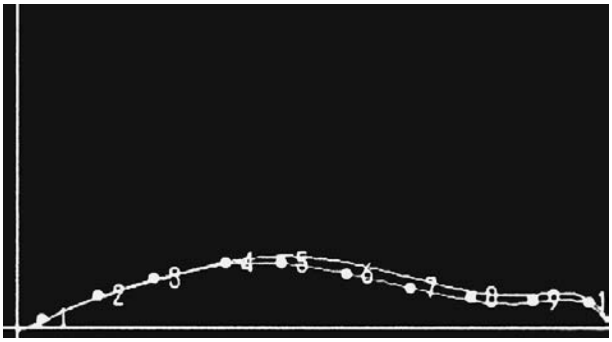


Figure 10 Polynomial function curves obtained from the forehead line in Fig. 9 showing a good match and 10 measurement points for calculating the distance between both curves. Reproduced from Yoshino *et al*. (1997).

Further Reading

Cai D and Lan Y (1993) Standards for skull-to-photo superimposition. In: İşcan MY and Helmer PR (eds) *Forensic Analysis of the Skull Sciences*, pp. 171–181. New York: Wiley-Liss.

Delfino VP, Vacca E, Potente F, Lettini T and Colonna M (1993) Shape analysis morphometry in computer-aided skull identification via video superimposition. In: İşcan MY and Helmer PR (eds) *Forensic Analysis of the Skull Sciences*, pp. 131–159. New York: Wiley-Liss.

Grüner O (1993) Identification of skulls: A historical review and practical applications. In: İşcan MY and Helmer PR (eds) *Forensic Analysis of The Skull Sciences*, pp. 29–45. New York: Wiley-Liss.

İşcan MY and Loth SR (1997) The scope of forensic anthropology. In: Eckert WG (ed.) *Introduction to Forensic Sciences*, pp. 343–369. New York: CRC Press.

Krogman WM and İşcan MY (1986) *The Human Skeleton in Forensic Medicine*, 2nd edn, pp. 413–457. Springfield, IL: Thomas.

Lan Y (1992) Development and current status of skull–image superimposition: methodology and instrumentation. *Forensic Science Review* **4**:125–136.

Lan Y and Cai D (1993) Technical advances in skull-to-photo superimposition. In: İşcan MY and Helmer PR (eds) *Forensic Analysis of the Skull Sciences*, pp. 119–129. New York: Wiley-Liss.

Martin R and Knussmann R (1988) *Lehrbuch der Anthropologie*, 4th edn, vol. I. Stuttgart: Fischer.

Sekharan PC (1993) Positioning the skull for superimposition. In: İşcan MY and Helmer PR (eds) *Forensic Analysis of The Skull Sciences*, pp. 105–118. New York: Wiley-Liss.

Seta S and Yoshino M (1993) A combined apparatus for

photographic and video superimposition. In: İşcan MY and Helmer PR (eds) *Forensic Analysis of The Skull Sciences*, pp. 161–169. New York: Wiley-Liss.
Stewart TD (1979) *Essentials of Forensic Anthropology*, pp. 255–274. Springfield, IL: Thomas.
Webster WP, Murray WK, Brinkhous W and Hudson P (1986) Identification of human remains using photographic reconstruction. In: Reichs KJ (ed.) *Forensic Osteology*, pp. 256–289. Springfield, IL: Thomas.
Yoshino M and Seta S (1989) Personal identification of the human skull: superimposition and radiographic techniques. *Forensic Science Review* **1**:23–42.
Yoshino M, Imaizumi K, Miyasaka S and Seta S (1995) Evaluation of anatomical consistency in cranio-facial superimposition images. *Forensic Science International* **74**:125–134.
Yoshino M, Matsuda H, Kubota S, Imaizumi K, Miyasaka S and Seta S (1997) Computer-assisted skull identification system using video superimposition. *Forensic Science International* **90**:231–244.

FIBERS

Contents

Identification and Comparison
Recovery
Significance
Transfer and Persistence
Types

Identification and Comparison

R Palmer, The Forensic Science Service, Huntingdon, Cambs, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0516

Introduction

Since the perpetrator of a crime may transfer fibers from his/her clothing to a victim, or vice versa, the subsequent identification of fiber type(s) comprising a target garment and the comparison of these with any suspect fibers, is a potentially invaluable method of demonstrating associations between individuals and/or places. Where fibers have been transferred during the commission of a crime, but no suspect has been identified, the identification of fiber types can be of immense probative value.

The scheme of analysis and types of comparison performed on fibers within the forensic context is crucial, in that each level of comparison performed should provide a progressive degree of confidence in any matches found. The number and nature of the comparisons performed will therefore have a profound effect on the evidential value of the findings.

There are many different methods of analysis which can be employed in the identification and subsequent comparison of textile fibers. The selection of these should be made in order to provide the greatest degree of discrimination early on, thereby providing the best possibility of excluding material and making the best use of the analyst's time. The methods of analysis employed by the modern forensic science laboratory are normally nondestructive, as it is often the case that the fibers themselves become a court exhibit in their own right and/or may be scrutinized for the purposes of a second opinion.

It is not the purpose of this article to provide a comprehensive coverage of all the theoretical and practical considerations associated with these methods, but rather to provide an overview of the methods and practices currently employed (or at least available) in the modern forensic science laboratory where the identification and comparison of fibers is carried out.

Where further detailed information regarding the theoretical and practical aspects of these techniques is required, reference should be made to the supplied further reading list.

It should also be pointed out that although all the methods and techniques described are certainly desirable, it may be that due to the financial constraints and/or a lack of suitable expertise in a given laboratory, these may not all be available to the analyst. Where such constraints on resources are a factor, then the most discriminating techniques affordable, should clearly be those employed.

Identification

The types of fibers used in the manufacture of textiles can be broadly defined as naturally occurring (e.g. wool, cotton) or synthetic (e.g. nylon, acrylic). There are many methods of identification of the various fiber types, ranging from simple microscopic appearance to much more complicated instrumental methods. In the past, it has often been the case that destructive techniques, such as solubility tests, melting point analysis, flaming tests etc., were employed. Although undoubtedly useful (and relatively inexpensive), such tests have fallen out of favor in the modern laboratory as: (a) they are often cumbersome and messy; (b) they require relatively large sample sizes; and (c) they are destructive. Most modern methods of analysis have the advantage of combining high levels of discrimination with few (if any) of the disadvantages of the more 'classical' methods. It is these methods which will be of primary consideration.

Microscopy

White light (brightfield) microscopy This is the simplest yet the most useful method of fiber identification and comparison. Under simple white light high-power microscopy, some naturally occurring fibers can be readily and speedily identified. Cotton (the most commonly encountered vegetable fiber), for instance, is characterized by its highly convoluted appearance caused by the collapse of the lumen cell wall during drying. Other types of vegetable fibers such as those employed in cordage material (e.g. hemp, jute, sisal) are more problematical, often requiring a combination of microscopy and chemical/physical tests for conclusive identification. Wool, being a hair, is readily identified by the presence of scales. Synthetic fibers on the other hand, are more difficult to identify using white light microscopy alone, as they are often fairly featureless. In such situations, the differences in the chemistry between generic fiber types (e.g. nylon, acrylic, polyester etc.) may be exploited using polarized light microscopy, in particular, by measuring the fiber's birefringence.

Birefringence Since synthetic fibers are polymeric entities, they exhibit a pseudocrystalline structure and are anisotropic (i.e. possess different physical properties in different directions of the structure). The anisotropic behavior of synthetic fibers utilized by this technique is that of the difference in refractive index between the parallel axis ($n_{\|}$) and the perpendicular axis ($n_{\perp}$) of the fiber, producing interference colors corresponding to the degree of difference between these two refractive indices, when viewed under polarized light. This difference is known as retardation, or path difference (i.e. $n_{\|} - n_{\perp}$). Where $n_{\|\|} > n_{\perp}$, a fiber is said to possess a positive elongation and a negative elongation when $n_{\|} < n_{\perp}$. Since retardation is also a function of the diameter or thickness of the fiber, birefringence can be defined as:

$$\Delta n = (n_{\|} - n_{\perp})/\text{diameter}$$

In practice, the path difference is measured not by measuring $n_{\|}$ and $n_{\perp}$ directly, but by the use of a *tilting compensator* or a *quartz wedge*. These devices introduced into the polarized light path between the sample and the observer, produce a graduated effect opposite to the path difference of the fiber. Once the effect is opposite and equal to the path difference of the fiber, the interference colors are extinguished – this is known as the extinction point (when a fiber is very intensely dyed, it may be difficult to establish this). The point at which this occurs can be read directly from the tilting compensator and the path difference of the fiber (in nanometers) extrapolated from a calibration table. It is therefore a simple matter to determine the thickness of the fiber using a calibrated graticule, to complete the equation:

$$\Delta n = \text{Path difference (nm)/diameter}(\mu\text{m}) \times 1000$$

Given that not all synthetic fibers are round, an approximation of the thickness may be made by inferring the cross-sectional shape by viewing it longitudinally (optical sectioning) and taking the measurement between the appropriate points, as indicated in **Fig. 1**.

In circumstances where the fiber has a very low Δn (e.g. acetate fibers) or a negative elongation (e.g. acrylic), then what is known as a first order red tint may be introduced in place of the compensator or wedge. This produces an increase in the path difference sufficient to distinguish between these fibers. Triacetate fibers have a path difference of 0, and therefore exhibit no anisotropy.

Birefringence measurement is therefore a useful technique for determining the generic class of fiber.

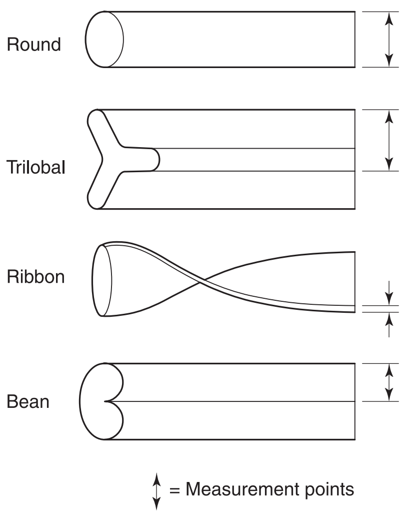


Figure 1 Typically encountered fiber cross-sections with fiber thickness measurement points indicated via optical (inferred) sectioning.

Typical birefringence values obtained from the various generic fiber types are shown in **Table 1**.

Since textile manufacturers often introduce copolymers in the production of a fiber to improve say, its fire-retardant properties, subtypes of fiber within a given generic class can be demonstrated. The technique of birefringence is not sensitive enough to distinguish between these. In such circumstances, an alternative method of identification/comparison is required.

Infrared spectroscopy

In any protocol designed for the identification and/or comparison of textile fibers, the technique of infrared spectroscopy is invaluable for identifying the polymers and copolymers employed in the production of a given fiber. Within the last decade or so, the availability of small, powerful, yet relatively inexpensive computers, has allowed this technique to evolve into Fourier transform infrared spectroscopy (FTIR) and such instruments have largely replaced dispersive instruments. It is beyond the scope of this article to describe the technical and theoretical issues concerning such instruments, however, they have the following advantages over the traditional dispersive instruments:

- They require minimal sample preparation
- They have a more favorable signal to noise ratio
- Higher precision
- Cleaner spectra
- Can be used on very small sample sizes.

Since this technique identifies the constituent polymers present in a fiber, the generic type of fiber can be readily identified. As already stated, textile manufacturers often employ copolymers within a process to alter some physical property of the fiber (e.g. flame retardation, increased tenacity, etc.) and so the identification of these chemical groups using FTIR allows a subclassification within a given generic type. In some circumstances it may be possible to use this information (perhaps combined with other information such as cross-sectional shape) to identify a particular brand name of textile and its manufacturer. Examples of spectra of two easily identifiable subclasses of acrylic fibers (acrylonitrile copolymerized with vinyl acetate and methyl acrylate respectively) are shown in **Fig. 2**.

Because of its ability to identify and subclassify synthetic fibers, FTIR can also provide a means of additional potential discrimination in any subsequent scheme of comparison.

Pyrolysis

Pyrolysis can be defined as the decomposition of a molecule by high temperature within a nonreactive atmosphere. This decomposition produces molecular fragments that are characteristic of the original

Table 1 Typical birefringence values obtained from different generic classes of fiber

Fiber type	Δn
Acrylic	−ve
Triacetate	0
Acetate	+ve
Viscose	0.025
Nylon	0.050
Polyester	0.195

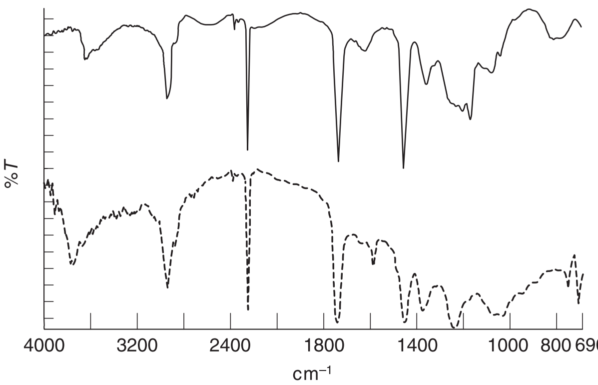


Figure 2 Examples of FTIR spectra of two subclasses of acrylic (polyacrylonitrile) fiber: acrylonitrile copolymerized with methyl acrylate (top) and acrylonitrile copolymerized with vinyl acetate (bottom). (Courtesy of The Police Forensic Science Laboratory, Dundee.)

material and can subsequently be analyzed and identified by techniques such as gas chromatography, mass spectrometry or FTIR. These techniques produce a pyrogram of the original material, which is used for identification and comparison purposes. Although this technique can potentially provide a high degree of discrimination, it is by nature destructive and requires relatively large sample sizes. Questions have also been raised about the reproducibility of the resulting pyrograms. For these reasons, this technique is not one which is universally employed in the identification and comparison of synthetic fibers, within the modern forensic science laboratory.

An example of a scheme for fiber identification using the methods outlined in this section, is given in **Fig. 3**.

Comparison

As with any comparison, care must be taken to ensure that the reference (control) samples are as representative of the target item as possible.

The comparison of suspect fibers with fibers comprising the target garment or item, begins with the searching of the surface debris tapings (or other methods of collection) using low-power stereomicroscopy. At this point, only general color and gross morphology is considered and therefore the search window is very wide. There are at the time of writing, various computerized automated systems designed to search surface debris tapings (e.g. Foster & Freeman FX5) and although these have certain limitations, they have been shown to be very useful time-saving tools. Any fibers not excluded by such a search (visual or automated) are removed and mounted using suitable media and coverslip onto a microscope slide for a more detailed comparison. There are numerous suitable types of mounting media available, and the choice of which to use should consider aspects such as the ease of removal of the fiber (for nonmicroscopic comparison), the stability of background color over time, and the nonreactivity with the fiber and its dye. The removal of a fiber from a microscope slide for subsequent analysis is usually achieved by breaking/cracking the coverslip and using a suitable solvent to liberate the fiber from the mountant. The fiber should then be thoroughly washed prior to analysis/comparison. Where a solvent-based mountant is used, care should be taken that on removal and subsequent washing of the fiber, leaching of the dye does not occur.

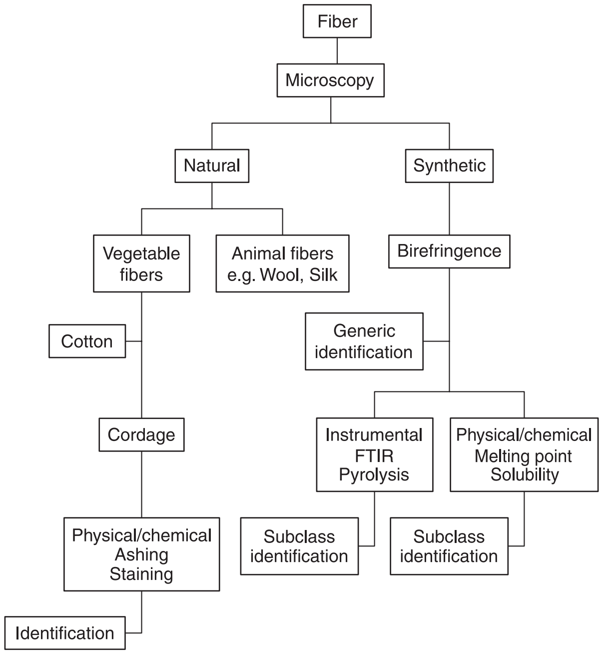


Figure 3 Flow chart showing scheme of fiber identification.

In any scheme of analysis/comparison, it is important that the most rapid, discriminating techniques are performed first, as this will serve to maximize the time efficiency of the examiner by quickly identifying exactly which fibers are required for further comparison and which can be excluded.

Such a scheme of analysis/comparison can be represented in flow chart form as in **Fig. 4**.

Where the size of a given fiber is optimal, the maximum number of tests possible should be carried out. Where this is not possible due to the limiting factor of the size of a single fiber, it may be more appropriate to carry out each of these tests on a selection of different fibers. Clearly, the more comparative tests that are performed on a given fiber, the greater the confidence of any match. In situations where the examiner is confronted by say, dozens of good microscopically matching fibers, then it may be acceptable to select a random sample of these for further testing.

Microscopic appearance

Without doubt, the most widely used piece of equipment in the forensic laboratory is the comparison microscope. In its basic form, this consists of two identical optically balanced microscopes, which have been linked via an optical bridge, in order to allow the simultaneous viewing of two separate samples under white light. Features such as the presence or absence of delustrant, inferred or direct observation of cross-sectional shape and color, can be determined and directly compared. By introducing polarized plates within the light path, direct comparison of birefringence interference colors can also be carried out. The

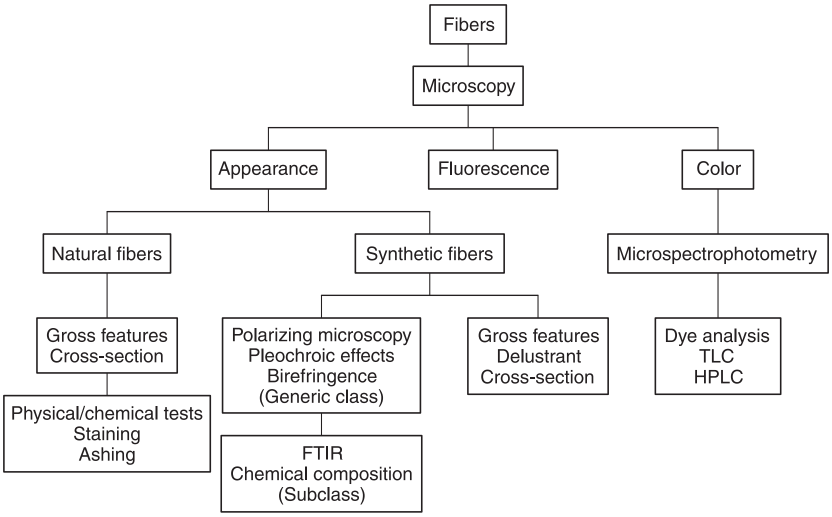


Figure 4 Flow chart showing scheme of fiber analysis/comparison.

direct visualization of the cross-section of a fiber (as opposed to inferred) can be achieved using various sectioning techniques and can provide useful comparative information, particularly where the modification ratio of a lobed fiber is determined. Briefly, the modification ratio can be defined as the ratio between the diameter of a circle subscribed by the outer tips of a lobed fiber and the inner circular core. Such measurements in addition to providing comparative information, can, when combined with other information (e.g. infra red spectra), be used to identify a specific brand of fiber.

Direct cross-sectioning is also extremely useful in the identification and comparison of plant fibers used in cordage materials.

Many (most) modern comparison microscopes possess various accessories such as specialist light sources and filter systems which allow the fluorescent characteristics of fibers to be compared using differing wavelengths of light. This is particularly useful in detecting the presence of optical brighteners.

Color comparison

The comparison of color using the comparison microscope is undertaken after ensuring that each microscope is correctly adjusted via Kohler illumination; the background color is neutral and balanced; the intensities are the same and that the same magnification is employed. Pleochroic (or diochroic) effects, that is the apparent differences in fiber color when viewed at different orientations under plane polarized light can also be compared.

Considerable intrasample variation (both quantitative and qualitative) in the color of fibers from a given garment may exist due to differences in dye intensity and dye uptake, especially where a multicomponent dye has been employed. This type of intrasample variation is predominately seen in naturally occurring fibers such as cotton and wool and is a function of the individual fiber's affinity with a particular dye. Similar variation can also be observed in synthetic fibers, where part of a garment has been disproportionately worn, bleached or exposed to sunlight. It is therefore crucial that a reference sample taken from a target garment should be as representative as possible, taking features of wear and tear into consideration. If the reference sample is not truly representative, the chance of false exclusions are increased.

Although the human eye can discriminate between subtle differences in color, the color of two fibers can appear identical under one illuminant condition, but different under another. This is known as *metamerism* and because of this, and the fact that color comparison is essentially subjective, this aspect of fiber comparison is augmented by the use of objective analytical methods.

Microspectrophotometry (MSP) As the name suggests, this is an adaptation of a standard visual range spectrophotometric analysis whereby spectra from small analytes can be obtained. In practical terms, such instruments consist of a microscope integrated into the light path of a single-beam configuration spectrophotometer. The detector of the system is linked to a computer for data handling and instrumental control. Some instruments work exclusively in the visible light range, whereas others such as the Zeiss MPM 800 series are capable of performing within the UV range (see later).

For visible range MSP, no special preparation of the fiber is required. In order to avoid any pleochroic

effects, care should be taken to orientate the reference and suspect fibers the same way for analysis.

MSP provides two types of comparative information: spectral comparison and color matching.

Spectral comparison The color of a fiber is determined by which parts of the visible electromagnetic spectrum are absorbed by its dye. A blue dye for example, absorbs predominantly the red part of the spectrum. A plot of relative absorbance against wavelength can therefore produce a graph of the absorption spectrum of a particular color. In the practical comparison situation, the spectra obtained from suspect fibers are compared against the spectral range observed in the fibers from the target item which can be performed by overlaying the spectral traces from the reference and suspect fibers on a light box or similar item, and determining the closeness of fit. Most computer data acquisition software such as the Zeiss 'Lambdascan', allow direct overlaying of spectra on the computer screen, often with the ability to 'zoom in' on specific parts of the spectrum under interest. Such displays can also be printed (**Fig. 5**) for hard copy inclusion in a case file. Where the spectrum of the suspect fiber falls within the qualitative and quantitative range observed in the reference sample, the fibers are said to match (i.e. are indistinguishable) in terms of color.

As with the microscopic visual comparison of color, care must be taken to ensure that the reference sample is as representative of the target garment as possible.

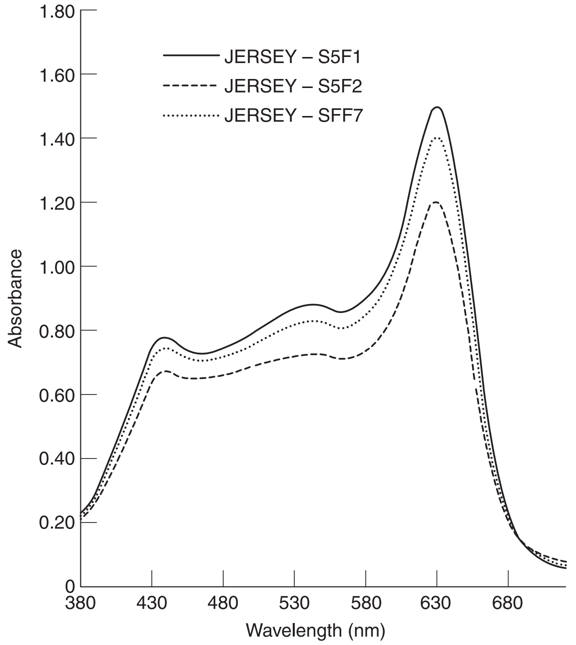


Figure 5 Microspectrophotmetry output showing overlaid spectra of colored fibers. (Courtesy of The Police Forensic Science Laboratory, Dundee.)

Color matching In addition to comparing the spectra of fibers directly, the spectral information can be used by the computer software to generate color data, that is to describe the color within the parameters of a particular mathematical model. This can be useful as a secondary means of color comparison and also as a means of databasing fiber color/type combinations for either estimation of prevalence of color/ type combination, or for intelligence purposes in an ongoing large inquiry.

Perhaps the most widely used model employed in calculating such color data is the C.I.E. (Commission International Eclairage) system, which is based upon the Young–Helmholtz theory of trichromatic color vision. Essentially, this model describes a color in terms of how much of each of the three primary colors of light (red, green and blue) are absorbed by the dye. Each of the primary colors is present in equal quantities in white light. The value of red is denoted as X, green as Y and blue as Z. These values are known as tristimulus values and hence, in the case of white light:

$$X + Y + Z = 1$$

These values are calculated using transmission data.

These values can vary according to viewing conditions and type of illuminant used. In the case of the CIE model, the conditions are defined as 'Illuminant C' – equal to that experienced on a bright overcast day.

Since tristimulus values have been shown to be nonlinear due to their derivation from transmission values, and textile fibers vary in width and dye uptake, different values for these can be obtained from fibers of the same color. The way around this is to calculate these values using absorbance data (=log(100/*T*)) and to normalize them:

$$x' = (X/X + Y + Z)$$
$$y' = (Y/X + Y + Z)$$
$$z' = (Z/X + Y + Z)$$

Since the wavelength of light transmitted is equivalent but not identical to the color absorbed, the color observed is complementary to the color absorbed. Hence these values are known as *complementary color coordinates*. By using these values, the color of a fiber can be defined and compared, ignoring the perceived effect of intensity.

UV microspectrophotometry In addition to visible range MSP, some forensic science laboratories are now using instruments capable of operating in the UV

range, producing absorbance and emission (fluorescence) spectra. Such instruments are produced by companies such as Zeiss and Nanometrics, and as expected in this type of analysis, have specialized optics and light sources. The increased discrimination afforded by these instruments has been shown to be useful in potentially distinguishing between fibers which had otherwise been indistinguishable using visible MSP alone.

Dye comparison

Thin layer chromatography (TLC) Where a fiber is too opaque for meaningful MSP analysis, or as an additional comparative technique providing a further degree of discrimination, TLC can be carried out on dye extracts from textile fibers. Given that the chemistry of the fiber dye interaction can be complex, certain classes of dyes (based on their chemistry) are used with particular fiber types (basic dyes for instance, are often used to color acrylic fibers).

The extraction of a dye from a fiber is usually achieved by placing the fiber in a sealed labeled capillary tube along with a suitable extractant. The extractant tube is then heated. Since this is a semi-destructive treatment, a portion of the fiber should be preserved on the microscope slide for any future scrutiny. The choice of extractant solvent depends on the nature or class of the dye present as these can have very varied solubility characteristics. Over the years, various sequential extraction schemes have been developed which not only provide the means of determining the best extractant solvent for a particular fiber/dye combination, but also determine the class of dye employed. Once the class of dye has been determined, the optimum eluent system can be used. The best choice of eluents for a given dye class have been determined by various studies performed over the years. Ideally an eluent should provide good separation of well-defined dye components (it is beyond the scope of this article to list the various extractant and eluent systems for the various fiber type/dye class combinations. This information can be gained from the supplied bibliography).

Once extracted, the dye is spotted onto a silica plate and placed in a chamber with the appropriate eluent. Components of the dye are selectively adsorbed as the eluent is drawn over the plate causing the dye components to be separated out. The solvent front of the eluted plate should be marked and the plate itself quickly dried. If appropriate standards and controls are applied to the same plate as the questioned material, a direct means of visual comparison of the dye components can be carried out. The retardation factor (R_F) values for each of the separated components can also be calculated. The eluted plate should also be visualized under UV light, in order to determine if any fluorescent component bands are present. Since the success of visualizing all the components of a dye depends on the amount of material extracted and applied to the plate, care should be taken to ensure that broadly similar amounts of reference and questioned dyes are applied.

Where it is not possible to extract the dye from a particular fiber due to a high fiber/dye affinity (e.g. wool/reactive dye) then a similar behavior with a questioned fiber, although by no means as conclusive as the production of a chromatogram, may indicate a common origin and serve as a means of excluding nonmatching fibers.

High performance liquid chromatography (HPLC) The technique of HPLC has also been applied to the comparison of dyes from textile fibers. Although this technique is useful in terms of increased sensitivity, it does present some technical and practical difficulties when used for this application. The main difficulty is that since fiber dyes are extracted in organic solvents, these are not easily compatible with aqueous-based HPLC eluent columns. This means that it is difficult (if not impossible) to obtain a specific instrumental configuration that will be optimal for the separation of dye components for every dye class. The provision of an expensive instrument of dedicated configuration (and therefore limited use) is not cost effective, especially where any advantage gained in its use, over TLC, is questionable. Until these practical difficulties are overcome, it is bound to remain a technique not normally employed on a routine basis.

Chemical composition

FTIR Where the recovered fibers are of sufficient length, part of an individual fiber can be used for FTIR analysis. As with TLC, it is important to leave a portion of the fiber on the slide in order that it can be examined microscopically if required. As with all comparison techniques, it is important that the preparation of the reference and recovered samples for analysis should be the same. The fibers should be flattened prior to analysis as this minimizes variation in beam path length through the fiber, giving better spectra. Flattening can, however, cause minor variations in peak position and intensity and therefore the compared fibers should be identically treated.

Once flattened, the fibers can be presented to the instrument by mounting them across a suitable aperture or IR window. Each sample should be aligned in the same direction in order to avoid any polarization bias effects which can occur with these instruments.

Sample and background scans should be run under identical conditions.

Most FTIR instruments use data management packages which not only display individual spectra, but can also overlay spectra for direct visual comparison purposes.

In addition, such software packages usually have 'tools' for the manipulation of the spectral data, specifically for functions such as noise reduction, normalization etc. It is important that, if these functions are used, the original data is saved in a separate file from the 'treated' data. As well as displaying and manipulating spectra many such software packages have a library of reference spectra with which questioned spectra can be compared for rapid identification. These also allow the analyst to add further spectra to the library and so over time, a comprehensive reference collection can be built up.

Chemical/physical tests In situations where infrared spectroscopy yields limited information in the identification and subsequent comparison of very closely chemically related fiber types (e.g. acetate/ triacetate, polyolefins), there may be little choice but to use destructive tests such as melting point analysis and/ or solubility tests. Clearly, in circumstances where these tests are employed, it is important that these are performed on the reference material first and that as little of the questioned material is used as is practically possible.

Documentation

As with any analysis, it is important that all the work relating to the identification and comparison of fibers is properly documented. All fibers should be given a unique number/slide allocation and the number and types of analysis performed on a given fiber, as well as the results, should be recorded. The record should also reflect the number of fibers recovered from a questioned source and the number of matches. In providing a clearly auditable record of work, the task of summarizing the work and interpreting the results is made easier.

In addition, there should also be a record of the routine maintenance and calibration of all analytical equipment used in the work.

Resources

As already mentioned, the given resources available to a particular laboratory and examiner in this field, may preclude the use of some or all of the methods outlined in this article. Where the identification and comparison of fiber trace evidence is to be practiced within financial constraints, then clearly the minimum requirements as opposed to the 'best practice' in methodology should be employed. It is the author's view that the 'minimum requirements' for this type of analysis should be the microscopic techniques described, combined with the 'classical' techniques (e.g. solubility testing) where appropriate.

Whatever resources are available to the analyst, the same methodical approach to this field of analysis should always be adopted.

See also: **Analytical Techniques**: Microscopy; Spectroscopy: Basic Principles. **Evidence**: Classification. **Fibers**: Types; Transfer and Persistence; Recovery; Significance.

Further Reading

Appleyard HM (1978) *Guide to the Identification of Animal Fibres*, 2nd edn. Leeds: WIRA.

Catling DL and Grayon J (1982) *Identification of Vegetable Fibres*. London: Chapman and Hall.

Challinar JM (1992) Fibre identification by pyrolysis techniques. In: Robertson J (ed.) *Forensic Examination of Fibres*. New York: Ellis Horwood.

Chamberlin GJ and Chamberlin DG (1980) *Colour – Its Measurement, Computation and Application*. London: Heyden.

Deforest PR (1982) Foundations of forensic microscopy. In: Saferstein R (ed.) *Forensic Science Handbook*. Englewood Cliffs: Prentice Hall.

Ferraro JR and Basile LJ (1979) *Fourier Transform Infra Red Spectroscopy Applications to Chemical Systems*, vol. 2. New York: Academic Press.

Gohl EPG and Vilensky LD (1985) *Textile Science – an Explanation of Fiber Properties*, 2nd edn. Melbourne: Longman Cheshire.

Grieve MC (1995) Another look at the classification of acrylic fibres using FTIR microscopy. *Science and Justice* **35**:179–190.

Grieve MC and Cabirers LR (1985) The recognition and identification of modified acrylic fibres. *Forensic Science International* **20**:121–146.

Hamilton R and Hamilton S (1987) *Thin Layer Chromatography*. Chichester: Wiley.

Hartshorne A, Wild F and Babb N (1991) The discrimination of cellulose di and tri acetate fibres by solvent test and melting point determination. *Journal of the Forensic Science Society* **31**:457–461.

Johri MC and Jatar DP (1979) Identification of some synthetic fibres by their birefringence. *Journal of Forensic Sciences* **24**:692–697.

Robertson J (ed.) (1999) *The Forensic Examination of Fibers*, 2nd edn. London: Taylor and Francis.

Stoeffler SF (1996) A flowchart system for the identification of common synthetic fibres by polarised light microscopy. *Journal of Forensic Sciences* **41**:297–299.

Technical Working Group for Materials Analysis (1997) *Forensic Fiber Examination Guidelines*. Washington, DC: Federal Bureau of Investigation.

The Textile Institute (1985) *The Identification of Textile Materials*, 7th edn. Manchester: The Textile Institute.
Tungol MW, Bartick EG and Mortimer A (1990) The development of a spectral database for the identification of fibres using infrared microscopy. *Applied Spectroscopy* **44**:543–549.

Recovery

R Palmer, The Forensic Science Service, Huntingdon, Cambridgeshire, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0515

Introduction

This article provides an overview of the methods available and the approaches used for the recovery and preservation of fiber evidence from items submitted to the laboratory and those present at a crime scene.

Since it is known that fibers can be readily transferred, the very process which makes them a potentially valuable form of trace evidence can also work against the examiner in the form of inadvertent 'contamination'. 'Contamination' in this context, refers to the introduction of material onto a questioned item, *after* an alleged contact has taken place. Thus, although any method of recovery must by definition be efficient at collecting fibers, it is crucial that it must also be equally efficient at preserving this evidence – not only in preventing its loss, but in preventing the introduction of 'contaminant' fibers. It is, therefore, of utmost importance that the methods and approaches used in the recovery of fiber trace evidence meet these requirements and are performed correctly – first time.

The methods and approaches to the recovery of fiber evidence outlined in this section should enable the reader to gain an appreciation of the practical issues and problems faced by the forensic fiber examiner in this primary, but crucial aspect of fiber examination.

Methods of Recovery

There are various methods of fiber recovery available, the choice of which may be determined by the circumstances of the case, condition and nature of the exhibit and personal preference of the examiner. Whatever methods of recovery are employed, they should be;

- Simple (easy to use);
- Rapid (time efficient);
- Efficient (optimal recovery);
- Preserve the evidence (prevent contamination);
- Allow easy subsequent searching (time efficient).

Arguably, some of the methods available may fit these criteria better than others. The relative merits and disadvantages of each are considered here.

Visual search

This method is clearly the simplest method of fiber recovery. A visual search of an item may often reveal transferred textile material, whether in the form of tufts or single fibers. Such a search can often be performed using a simple hand lens and is well suited to situations, such as burglary, where the perpetrator has left fibers at a point of entry, such as a broken window. This can also be considered when dealing with weapons such as knives, or on clothing from a body which has been transported in an item, such as a carpet and subsequently dumped.

Such material can be removed using forceps and then placed in a sealed, labeled container, pending further examination. The advantage of this approach is that the recovered fibers can be rapidly removed from the container for subsequent examination without having to be searched for amidst 'background' debris. This method is particularly useful where the item in question (such as a shoe) which although poorly retentive in itself, may have fibers trapped in crevices or on some adhering medium, such as chewing gum. The detection and recovery of fibers in such circumstances can be greatly aided by the use of low power stereomicroscopy. This method should be considered before any other is attempted.

Since fibers are mainly transferred in the form single tiny fragments, rather than visible tufts or threads and are therefore not usually obvious to the unaided eye, other methods of recovery and searching of general fiber debris from an item have to be employed.

Surface debris tapings

Extraneous fibers can be recovered from the surface of a garment or other item by the use of transparent adhesive tape. This is achieved by taking a length of such tape, holding it adhesive side down and systematically dabbing the surface of the item in question (**Fig. 1**). The tape is then stuck down onto a clear plastic sheet, effectively preserving the recovered debris. This process is repeated until all of the surfaces in question have been covered. Completed tapings are then clearly labeled and placed in an appropriately labeled greaseproof envelope or similar receptacle. As an alternative, the tape can be attached to a 'roller'

Figure 1 (see color plate 21) Recovering fiber debris from a garment using clear adhesive tape.

(available from hardware stores and normally used to press down wallpaper edges) which is systematically rolled across the garment, replacing the tape as necessary. This method allows a greater degree of control of the pressure applied to the taping and hence the efficiency of recovery.

As well as providing an efficient means of preserving the collected fiber debris, surface debris taping also allows easy subsequent searching using low power stereomicroscopy. The area of tape around any potential fiber match is marked and a 'window' cut around the fiber in question using a scalpel (**Fig. 2**). Using an appropriate solvent, the fiber is removed and mounted directly onto a microscope slide using a suitable mounting medium (**Figs 3** and **4**). The microscope slide is appropriately labeled and placed in a covered slide tray, again which has been appropriately labeled.

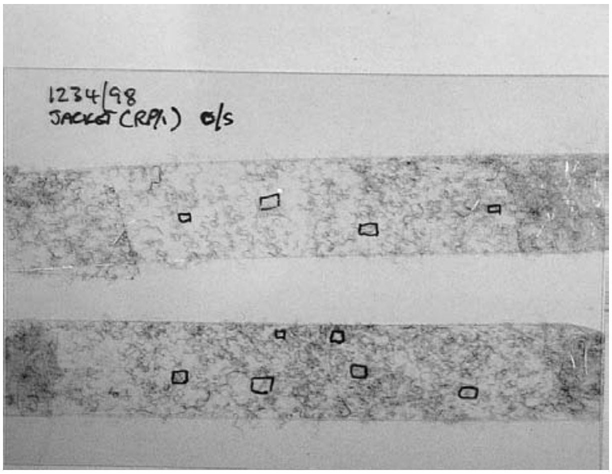


Figure 2 Tapings bearing fiber debris stuck onto acetate sheet. Note marked areas indicating potential matches.

Figure 3 (see color plate 22) Searching for and removing matching fibers from surface debris tapings using stereomicroscopy.

Care should be taken to prevent overloading the tape, i.e. picking up too many 'background' fibers ('background' fibers refer to fibers which are part of the construction of the recipient garment, or any

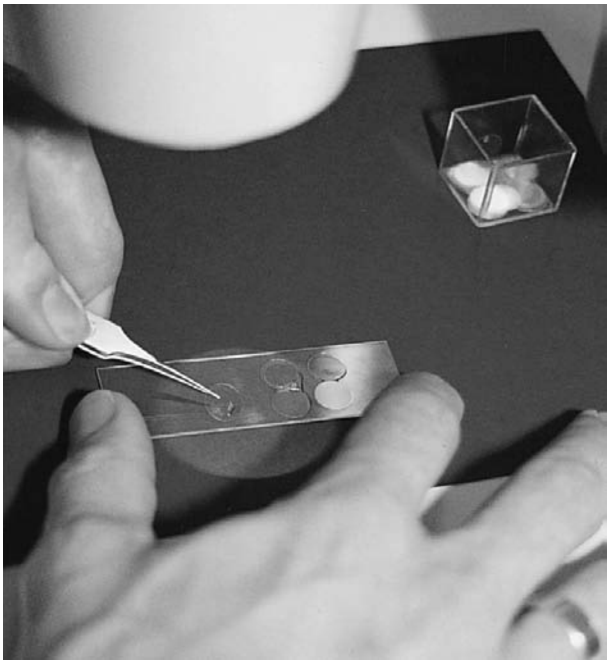


Figure 4 (see color plate 23) Mounting recovered fibers onto a microscope slide.

other extraneous fibers present on the item's surface which are not targeted). In addition to making it more difficult to search (as one has to search through all the 'background' fibers), this also reduces the adhesive ability of the tape and therefore seriously compromises the efficiency of recovery. It is, therefore, better to use many tapes on a given item, as this will optimize the efficiency of recovery by reducing 'overloading' and in turn, aid subsequent searching. This problem can also be alleviated to a certain extent by using a low-adhesive tape on items which readily shed their constituent fibers, or where a high applied pressure of taping using 'rollers' is employed.

The taping method of fiber recovery also lends itself to the use of computerized automated searching instrumentation, of which there are several available on the market. Although the use of surface debris tapings is the method of recovery most frequently employed by forensic scientists, it does however have certain drawbacks;

- It is unsuitable for use on wet or heavily contaminated items, or items where there are large amounts of particulate debris.
- It can be difficult to use on angular or irregular surfaces.
- It does not easily lend itself to use on large surface areas.

Scraping

This technique involves suspending the item in question above a collection funnel or sheet of paper and scraping the debris from the item onto the recipient item. The debris is subsequently transferred to a Petri dish or similar receptacle. Where a collection funnel is used, the debris falls directly into the receptacle. This method of collection can be useful for items which are soiled, or where several types of trace evidence are to be recovered simultaneously. However, the absence of published scientific data relating to concerns over the efficiency of this method and the potential for contamination, have meant that this method is not universally regarded as best suited for the recovery of fiber evidence.

Vacuuming

This technique uses a modified vacuum cleaner with a collection filter incorporated into the hose system. This method is useful for items with large surface areas such as carpets, and items which are soiled with particulate debris, such as car footwells. The main disadvantages with this method are that the apparatus must be scrupulously cleaned to prevent contamination and the fact that it tends to be (ironically) too efficient, in that it indiscriminately recovers large masses of debris making searching for target fibers difficult. The efficiency of this method of collection also varies between different machines.

Combing

This method uses a 'seeded' comb, i.e. a hair comb to which cotton wool has been introduced between the teeth (**Fig. 5**). This is used primarily to recover extraneous fibers from hair in cases of armed robbery or terrorism where a fabric mask has been worn over the perpetrators head (**Fig. 6**). The introduction of cotton wool between the teeth of the comb increases the efficiency of collection by introducing a highly retentive collection surface to the recovery system. Studies

Figure 5 (see color plate 24) The seeded comb.

Figure 6 Use of seeded comb to recover fibers from head hair.

have shown that such fibers transferred from such garments can persist on the head hair for considerable periods of time, even after washing. This method can also be applied in cases of rape, where it is suspected that fibers may have been transferred to the pubis of the victim or assailant.

Once the comb has been passed through the suspect's hair, the cotton wool (and the comb) are examined for the presence of fibers using stereomicroscopy. Since cotton wool is (generally) white, this also aids detection of transferred fibers. Any such fibers found are removed directly and mounted onto a microscope slide (**Fig. 7**).

Care must of course be taken to ensure that the comb and cotton wool used are free from contaminants. The comb should be provided in a sealed tamper evident bag and not used if the packaging appears compromised.

Choice of method

In considering the various methods of recovery of fiber evidence, it is important to distinguish between a laboratory environment and that encountered at the crime scene. Whereas the former is carried out in strictly controlled conditions with access to microscopes, good illumination and other instrumentation, the latter is often performed under difficult circumstances, where the environment is unpredictable, dirty and poorly illuminated, with limited access to specialist equipment. Since many crime scenes are outdoors, weather conditions may have caused much of any trace evidence to have been lost, or indeed in the case of rain and wet surfaces, render methods of collection such as surface debris taping useless. In such circumstance the analyst may be confronted with what is essentially a damage limitation exercise, using whatever methods are appropriate to the conditions – even though they may not be (under ideal conditions) the most efficient.

Figure 7 (see color plate 25) Searching for recovered fibers on the seeded comb.

Given a crime scene indoors and sheltered, the opportunity presents itself to recover fibers immediately, minimizing any potential losses and contamination. In cases of murder, it is appropriate to take surface debris tapings from the clothing and skin of the deceased *in situ*. A close visual search of the body (particularly the hands) prior to taping, may also reveal fibers which can be removed by forceps. This approach minimizes any potential losses and/or contamination that may occur as a result of the body being removed from the scene to the morgue. In addition, so called 'one to one' taping may be employed. This is a method whereby surface debris tapings are placed over the exposed surfaces of the body and clothing. The position of each of these tapings is catalogued. Any recovered fibers subsequently recovered from these tapings can therefore be associated with a particular location on the deceased's body, allowing a 'distribution map' of any transferred fibers to be built up. This may provide useful additional evidence in certain circumstances. Since clothing removed from an individual is packed in a bag prior to submission to the lab, fibers may be redistributed from one area of the garment to another during transit, thus rendering this 'one to one' approach inapplicable at the laboratory.

It can be seen then, that although the crime scene can present some technical difficulties in the recovery of fiber evidence, it can in certain circumstances, present opportunities to maximize recovery and provide additional information.

Although it may be the case that one particular method of fiber recovery is routinely employed by a particular laboratory, it may be that the circumstances of a particular case (whether at the crime scene or in the lab) dictate that a method other than the particular 'norm' is more appropriate. It is important that each of the above methods of recovery are not seen in isolation, as it is not unusual that a combination of some or all of the above methods becomes appropriate in certain circumstances. Given that the object is to recover and preserve fiber evidence, there is nothing to be gained and literally everything to be lost through blinkered thinking and an inflexible approach.

Documentation and Packaging

Whichever method of fiber recovery is employed, it is imperative that the recovered material is preserved in a manner which prevents loss and contamination

until such times that it can be examined. Since it is often the case that material is recovered by an individual other than the analyst, it is vital that any such material is clearly and unambiguously labeled as this information may become vital in demonstrating to the court the continuity, integrity and significance of its recovery. Even where the analyst is the person recovering the material, the same scrupulous approach must be taken for the same reasons. Since it is not unusual for fiber debris to be recovered from many items in a given investigation, then a meticulous approach to this aspect of documentation will have the added benefit that it is easier for the analyst to develop an examination strategy when it is known exactly what there is to work with.

Labeling

All fiber debris, whether recovered at the laboratory, or crime scene, should be comprehensively labeled, indicating:

- The item from which it was recovered and the location of that item;
- The area on the item from which it was recovered;
- The date (and time) it was recovered;
- The person(s) who recovered it and those subsequently taking receipt;
- An alphanumeric identifier unique to a particular item.

The label of the recovered debris should be signed by everyone who has been involved in its chain of custody, examination or collection and this should be securely fastened to its packaging.

Packaging

The debris should be packaged using a suitable receptacle (ranging from a simple envelope to a brown paper sack) which should be securely sealed, preferably with some form of tamper evident system. Such a system can simply be achieved by a signed label stuck underneath clear sealing tape. The label, bearing the relevant information regarding the sample, should be securely attached to the outside of the packaging. This, of course, also applies to clothing and other items.

Documentation

Where fiber debris is recovered and packaged at a scene, details should be recorded in a register by a designated exhibits officer. Since the significance of material recovered from a scene of a crime may not become evident for many weeks or even years after a crime, it is vital to be able to establish what was actually recovered, from where and by whom. Details of where and when a particular item was stored prior to its examination should also be recorded. The meticulous cataloging of a recovered exhibit can make the difference between a breakthrough being made in a long running investigation, or the investigation coming to a 'dead end' due to a crucial item being left on a shelf in a store somewhere, with its existence and significance long forgotten.

On receipt at the laboratory, the condition of the packaging and labeling should be noted and any problems likely to compromise any potential evidence should result in the item in question being rejected for examination. The examiner should also document the location, time and date at which the item was examined and a detailed description of the item and recovered debris should be made in his/her notes. Again, this is crucial in demonstrating to the court the continuity and integrity of any evidentially significant findings.

Contamination Issues

Since fibers can be readily transferred, care must be taken to insure that this does not work against the examiner in the form of inadvertent contamination. 'Contamination' in this context refers to contact or fiber transfer between two items which are alleged to have an association, *after* the alleged association has occurred. Clearly if contamination can be demonstrated, then any subsequent demonstration of fiber transfer between the items in question will have no evidential value. In addition to preventing contamination by keeping questioned items separate, care must also be taken to prevent the *secondary* transfer of fibers between items. An example of the secondary transfer of fibers might be when a victim of a crime has been transported in a police car which is subsequently used to transport a suspect. In such a situation there is a distinct possibility that fibers from the victim could be transferred to the car seat and these in turn transferred to a suspect's clothing.

Case history 1: The police were called to an old people's home where an intruder had sexually assaulted one of the female residents. Two police officers attended, one remaining in the car, while the other entered the building and room of the victim to take details. The officer remaining in the car, saw a man acting suspiciously nearby and confronted him. Unable to give any credible explanation for his presence there, he was held by the police officer until his colleague returned. On his colleague's return, both officers physically searched and arrested the suspect. The suspect was placed in the patrol car, and the police officer returned to the victim's room and packaged the bedding in a bag. On returning to the

car with the bag containing the bedding, the two police officers, suspect and exhibit, proceeded to the police station. The bedding from the victim's room and clothing from the accused were submitted to the forensic laboratory for examination. The laboratory found four fibers on the bedding which could have originated from the accused's clothing.

The suspect was eventually acquitted at trial, as it was successfully argued that the fiber findings in this case could have been due to crosscontamination by the arresting police officers.

Similar, secondary transfer could occur by placing two items of clothing on the same bench (even at different times) for packaging by the same person.

Case history 2: A young girl had been brutally raped by an unknown assailant and was conveyed to a medical examination facility for intimate medical samples to be taken. These as well as her clothes were submitted to the forensic laboratory for analysis. A large number of red acrylic and wool fibers were found on the surface debris tapings taken from her clothing. Clearly such findings were of probative value to the subsequent investigation and formed a large part of the police inquiry effort. After many weeks of investigation, it was found that these fibers matched that of a blanket in the medical examination suite. It was eventually established that this blanket had been placed around the victim by a well-meaning nurse, prior to the victim's examination.

It can therefore be seen that it is of the utmost importance to implement measures to prevent contamination before the recovery of fibers begins. The following are examples of such measures:

- Victim and accused should be transported in separate cars;
- Victim and accused should be interviewed and/or examined in separate rooms;
- Clothing should be packaged in separate rooms by different individuals;
- Protective clothing (such as paper boiler suits) should be used at crime scenes to avoid the introduction of fiber material to the scene and/or to prevent transfer of material to items potentially associated with that particular scene.

Once at the laboratory care must again be taken by the examiner to prevent primary or secondary fiber transfer between questioned items before any potential fiber evidence has been recovered and preserved. Methods to prevent such contamination at the laboratory:

- The integrity of the labeling and packaging of items received at the laboratory should be checked;
- Questioned items should be examined and material recovered from each in separate rooms, ideally in different parts of the laboratory;
- Instruments such as forceps, etc., lab coats and other items involved in the examination should be peculiar to each room and be left there after each examination. Adhesive tapes used in recovery should be kept packaged;
- Examination benches, collection funnels and vacuum equipment should be thoroughly cleaned before and after use. Examiners should also wash their hands before entering and leaving a room where fiber recovery is about to be, or has been performed;
- Since fibers are minute entities, care should be taken that any air handling/conditioning system does not potentially blow these away from or around the recovery area.

Conclusions

It can be seen that although methods of fiber recovery should be optimal in terms of the efficiency of recovery, it is absolutely crucial that the integrity of any recovered material is preserved through the proper approaches and applications of these methods. At best, a sloppy approach to this fundamental aspect of fiber examination can result in many hours of work being rendered inadmissible or worthless in court, or in the worse case scenario, an innocent person being implicated in a crime.

It cannot be overemphasized that the subsequent analysis and interpretation of fiber evidence is dependent on getting this relatively simple aspect of fiber examination correct the first time, as mistakes cannot be rectified later.

See also: **Crime-scene Investigation and Examination**: Recording; Recovery of Human Remains; Packaging; Preservation; Contamination. **Fibers**: Types; Transfer and Persistence; Recovery; Significance.

Further Reading

Fisher BAJ and Svensson W (1987) *Techniques of Crime Scene Investigation*, 4th edn. New York: Elsevier.

Kidd CBM and Robertson J (1982) The transfer of textile fibres during simulated contacts. *Journal of the Forensic Science Society* **22**:302–308.

Lowrie CN and Jackson G (1991) Recovery of transferred fibres. *Forensic Science International* **50**:111–119.

Lowrie CN and Jackson G (1994) Secondary transfer of fibres. *Forensic Science International* **64**:73–82.

Moore J, Jackson G and Firth M (1984) The movement of fibres between working areas as a result of routine examination of garments. *Journal of the Forensic Science Society* **24**:394.

Palmer R (1998) The retention and recovery of transferred fibres following the washing of recipient clothing. *Journal of Forensic Sciences* **43**:502–504.
Pounds CA (1975) The recovery of fibres from the surface of clothing for forensic examinations. *Journal of the Forensic Science Society* **15**:127–132.
Robertson J (1992) *The Forensic Examination of Fibers*. Chichester: Ellis Horwood.
Saferstein R (1988) *The Forensic Science Handbook*, vols 1 and 2. Englewood Cliffs, NJ: Prentice Hall.
Salter M and Cook R (1996) Transfer of fibres to head hair, their persistence and retrieval. *Forensic Science International* **81**:211–221.
Technical Working Group for Materials Analysis (1997) *Forensic Fiber Examination Guidelines*. Washington DC: Federal Bureau of Investigation.

Significance

C Roux, Department of Chemistry, Materials and Forensic Science, University of Technology, Sydney, Australia
J Robertson, Forensic Services, Australian Federal Police, Canberra, Australia

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0517

Introduction

When two or more fiber samples cannot be distinguished after thorough laboratory examination and consideration of similarities and dissimilarities, the question remains of what significance can be attached to this result. Sometimes case workers will express this finding in terms of a ‘could have come from’ statement. What is the value of this type of statement? This question constitutes probably the main challenge faced by the forensic fiber examiner. Unlike glass or paint evidence which may be used sometimes as ultimate proof when a physical fit can be demonstrated, fibers (excluding pieces of fabrics and some exceptional cases) are always considered as class evidence. However, this does not mean that fibers have little or no value. The evidential value of fibers depends on a multitude of parameters and is often stronger than one would expect. Fibers can even constitute primary evidence under some circumstances. It is the aim of this article to present and discuss the value of fiber evidence.

Factors Influencing the Significance

Fiber evidence is difficult to interpret. Although the fundamental parameters such as relevance of the trace, transfer/persistence and frequency in the population of interest are similar to those related to other trace evidence such as glass or paint, the problems associated with fibers are more difficult. For example, it is not possible to build and maintain a database of clothing types in the same way that one would do with a database of glass types. In most cases the questions raised in relation to the significance of fibers are extremely complex.

Since the mid-1980s different aspects of the evidential value of fibers have been studied, and there is now a clearer view on the significance of fiber evidence. It is generally accepted that the value of fiber evidence depends on many factors which are often complex and interacting. These factors can be classified as known (or usually determinable) or unknown (or difficult to determine).

Factors that are known (or that are usually determinable)

Circumstances of the case The discovery of fibers may be more or less significant depending on the circumstances. The forensic scientist needs to consider whether or not fibers can be explained by coincidence, by a real transfer but one which has an explanation which is not incriminating or the latter but where the explanation may be incriminating. In order to attempt to choose between these options the scientist needs to know relevant and often detailed information about the circumstances of the case and details relevant to the individuals allegedly involved.

Time elapsed before collection of the evidence The longer the time between the transfer event and the collection, the weaker the evidence. The risk that the relevant fibers are being lost and replaced by non-relevant fibers via secondary transfer is increased with time.

Suitability of the fiber types for recovery and comparison Some fibers are highly colored or luminescent, which means they are not only easier to find and collect, but also have more features to examine and to compare with each other. This increases the probability to discriminate fibers which do not have a common origin, and therefore enhance the significance.

Extent of the comparative information derived from the samples This information may be limited by the laboratory resources or by the size and type of fibers. It is obvious that a ‘fiber match’ is much more

significant when complementary and highly discriminating techniques have been used in comparison to simple screening tests.

Number of types of matching fibers Garments often contain more than one fiber type. This means that a 'multiple match' increases the significance of fibers. It should be noted that, in cases involving a blended fabric, the absence of one or more fiber types in the pool of recovered fibers does not necessarily preclude the blended fabric from being the source of the questioned fibers. Different components of a blended fabric may have different shed potential, and the number of fibers transferred of the different types are not necessarily proportional to the stated composition of the garment. This phenomenon is well known and called differential shedding.

Whether or not there has been an apparent cross transfer of fibers This situation arises where, for example, fibers found on the victim's garment are not differentiable from the fibers of the suspect's garment and fibers found on the suspect's garment are not differentiable from the fibers of the victim's garment. The demonstration of a cross-transfer constitutes the ideal situation following the Locard exchange principle. In this case, the significance of fiber evidence is dramatically increased, because the chance of finding matching fibers in both samples are remote.

Number of matching fibers recovered In general the larger the number of fibers the smaller the chance of finding these fibers by chance only. The discovery of a small or an unexpected number of fibers is difficult to interpret and requires good knowledge of transfer and persistence theories.

Location of the recovered fibers Some locations are more prone to secondary (nonrelevant) fiber transfer than others. In a break-in case, for example, fibers found at the edge of a smashed window are more significant than fibers found on the ground. The link between the perpetrator and the evidence in the latter case is quite unclear, which decreases the significance. Similarly, fibers found on undergarments in a sexual case may be more significant than fibers found on outer garments, for example.

Methods used to conduct the examinations More discriminating and complementary methods will bring more significant comparative features and will decrease the risk of coincidental match.

Factors that are unknown (or that are difficult to determine)

Extent and force of contact These factors influence fiber transfer and persistence with respect to both the number of fiber types and the number of fibers.

Degree of certainty that specific items were definitely in contact In some cases there is a higher degree of certainty that the items submitted are those involved in the alleged incident. For example, there may be good eyewitness accounts or the suspect was apprehended quickly. In other cases it is quite possible that the clothing is not relevant to the incident. For example, a suspect had the opportunity to dispose of the clothing, or eyewitness statements are unclear.

Donor fiber shed potential Some fabrics shed more fibers than others, and this must be considered when assessing the significance of the number of fiber types and the number of fibers. This means that it is impossible to give a simple rule which would define a cut-off number of fibers beyond which the primary transfer is certain, as opposed to secondary or higher transfers.

Frequency of occurrence of the matching fiber types The significance is weighted by the frequency of occurrence of the matching fibers. It is obvious that a match involving common fibers is less significant than that involving rare fibers. It should be noted that, owing to the high degree of variability of fibers, this frequency is generally much smaller than one might expect in the first instance.

Assessment of the Significance and Information Available

The assessment of the significance of a fiber finding requires a systematic study of the different factors described above in light of the circumstances of the case and the hypotheses alleged by the different parties (generally defense and prosecution). This process requires a great deal of experience and logical reasoning. Some valuable information exists through specialized literature or unpublished research projects, and can be used as an aid by the fiber expert during the interpretative process. However, in some cases, the lack of relevant information may prompt the set up of simulation experiments in order to confirm or deny some hypotheses which remain unresolved. Although interpretation problems have engendered great debates within the forensic community over the past 10–20 years, significant advances

have been made in interpretation modeling. Unfortunately, the practical use of the developed theories is limited as long as there is a lack of background data available. The importance of the problem requires fundamental research of high caliber to be undertaken through collaborative projects between academia and operational agencies.

Most of the information currently available in the area of the significance of fiber evidence is related to the following issues:

- fiber transfer and persistence;
- frequency of occurrence of fiber types; and
- statistical interpretation.

Fiber transfer and persistence

Knowledge of fiber transfer and persistence assists in answering the question of whether or not the number of fibers and the number of fiber types found in a given case fits the allegation of contact. In other words knowledge on transfer and persistence assists in answering the competitive questions 'what is the probability of finding the number of fibers and fiber types found in a given case if there was a contact?' and 'what is the probability of finding the number of fibers and fiber types found in a given case if there was no contact?'

This area of investigation includes the following topics:

- transfer studies;
- studies on differential shedding;
- experiments on the alteration of fiber characteristics in manner consistent with localized conditions specific to a case; and
- persistence studies.

Since 1975, numerous transfer and persistence studies involving fibers have been undertaken. The results give general guidance when assessing the number of fibers expected to be found under a given set of circumstances. It is beyond the scope of this chapter to examine these results in fine detail, but the most important findings are given below. The number of fibers to be found mostly depends on:

- the area of contact;
- the kind of physical contact (number of contacts, pressure, friction, time, etc.);
- the construction of the fabrics involved (donor and recipient);
- the fiber types (generic class, density, diameter, etc.);
- the length of the fiber fragments; and
- whether or not, and how, the recipient is mobile after the transfer event has occurred.

Small fiber fragments on the outer surface of a garment are more likely to be transferred and generally are more persistent on the recipient. Transferred fibers are more or less rapidly lost and redistributed on other locations (on the same recipient or not). Experimental studies have shown that fibers transferred at the time of contact may range from only a few to many hundreds or thousands. When the recipient is mobile, it has been shown that there is a rapid loss of fibers (typically 80% of loss within the first 4 h and 5–10% remaining after 24 h). On the other hand, fibers can persist for periods of many days and even weeks when transferred to a recipient that remains subsequently relatively undisturbed such as a car seat or a dead body.

As already stated, in blended fabrics the number of fibers transferred of the different types is not necessarily proportional to the stated composition of the garment. This issue, called differential shedding, is important when interpreting findings involving a fabric of blended composition.

The presence of only a few matching fibers may mean that they have been deposited by means of a secondary or subsequent transfer as opposed to a primary one (secondary transfer is the first indirect transfer after primary transfer, taking place via an intermediary object; common in contacts involving seating). Other reasons may also account for this situation, including a long time gap between contact/transfer and examination, a redistribution of fibers (e.g. due to the washing of the garment), the use of an inefficient method of fiber recovery, and a coincidental transfer. In all cases, caution is necessary when interpreting the finding of a small number of fibers especially to items such as underclothing.

In real life the variables which will contribute to the number of fibers that may be transferred are so numerous and unknown that the most reliable assessment of the number of fibers to be found is done by the simulation of the suspected contact with the actual fabrics in the case under investigation. However, pressure of casework in most laboratories would make this type of research work impossible in all but the most critical cases.

Frequency of fibers

Knowledge on the frequency of fibers helps the assessment of the relative rarity of the fiber features observed during the examination. In other words knowledge on the frequency of fibers assists in answering the competitive questions 'what is the probability of finding the fiber features observed in a given case if there was a contact?' and 'what is the probability of finding the fiber features observed in a given

case if there was no contact?' Although there is still a lot to learn on the frequency of fibers, much more information is available than was the case 10 years ago. This information comes from the following sources:

- population studies;
- target fiber studies;
- data collections; and
- trade inquiries.

The population studies have analyzed examples of the normal fiber population that may be expected on a given surface. This is vital information to have at hand when trying to answer the defense question 'well surely you would expect these types of fibers to be present on any such surface?'. Fiber population studies involve sampling the given surface and classifying the recovered fibers into fiber type and color categories.

Conclusions which can be drawn from the few studies which have been carried out so far have shown that:

- synthetic fibers form only a low percentage (13–20%) of any population yet studied;
- the majority of color/morphology/generic type combinations (approx. 65%) are represented by a single fiber type only. Synthetic fibers exhibit a very high degree of polymorphism;
- the chance of one type representing >1% of the total population is remote. A collection of synthetic fibers (same type, same morphology, same color) can therefore be said to be the result of a recent contact with a specific textile and can be considered highly significant

Target fiber studies give some idea of what proportion of the population has fibers of a particular type on their clothing (or car seat, cinema seat, pub seat, etc.). These data incidentally will automatically take into account not only the rarity of a fiber type, but also the tendency for this fiber to be transferred or to persist on the receiving fabric. So far, five target fiber studies have been carried out, involving:

- a blue wool nylon pullover from Marks & Spencer with sales over 1 million;
- red wool from ladies pullover and brown polyester of men trousers on car seats;
- green cotton fibers from Marks & Spencer's leggings and red acrylics from a pullover on car and cinema seats;
- red acrylics from a scarf sold over a five-year period in quantities of 5000–10 000 annually in nine European countries;
- blue wool fibers on seats in public houses throughout the United Kingdom.

These studies suggest that it is a rare event to find a fiber, even from a very common garment, on clothing (or a seat) without the donor recently having been in close contact with the garment in question.

Databases are generally built using garments received in forensic laboratories for examination. They yield very detailed information, such as the:

- frequency of morphological characters within a fiber type;
- frequency of a certain polymer composition within a fiber type;
- frequency of uncommon fiber types in the general population;
- frequency of usage of fiber types in different textiles divided into different categories;
- frequency of certain fiber type combinations in different textiles.

One of the criticisms of databases is that the collection of materials takes too long to accumulate and, hence, the data quickly becomes out of date. However, this criticism only really applies where the measurement of color data is involved. To overcome this problem, some authors have built up a comprehensive and easily updated database using data on garments offered for sale in mail order catalogues. However, this approach is feasible only in countries where mail order is popular.

Trade inquiries involve tedious and time-consuming investigative work. However, such inquiries can bring extremely valuable information on the relative frequency of a given type of fiber, including sales figures and manufacture figures such as quantity of items dyed using the same dye batch, for example.

Statistical treatment (probabilistic model)

The interpretative process is not entirely free from subjective opinion when dealing with fiber evidence because of the number of interactive factors which need to be considered. As a result, a probabilistic model has been proposed as an aid to this process.

Taking note of the fact that the decision process at the time of the trial is based on inference, it appears that the logical rules for thinking about facts in legal cases are those of probability. Only probabilities are able to measure the uncertainty that exists and evolves during every criminal case. So, only a model supported by the mathematical theories of probability is valid.

During the past 20 years, many researchers have shown increasing interest for a probabilistic model based on the Bayes Theorem. This model describes how the elements of proof combine with themselves and evolve during the decision process of the trial,

avoiding distortion based on inappropriate use of intuition. It is not the purpose of this article to develop Bayesian theory and the mathematical developments that result from it; suffice to say that when applied to fibers, as to other transfer trace evidence, Bayesian treatment can help to formulate the most relevant questions in the form of an expert system. The relevant questions are dependent on the circumstances of the case, but they are all derived from two fundamental questions:

- what is the probability of the evidence if the alleged story is true?
- what is the probability of the evidence if the alleged story is not true?

The answer, although subjective, can be combined with objective information to provide the most accurate guide to interpretation currently possible.

The main limitation of a probabilistic model is the fact that such a model uses data that are incomplete and difficult to quantify. For example, considering a case where k groups of matching fibers amongst n groups of fibers have been discovered, the Bayesian approach requires the assessment of the following probabilities:

- probability that k groups of fibers have been transferred and persisted;
- probability that there is normally $n-k$ groups of fibers on this kind of recipient;
- frequency of the matching fibers in the population of fibers present on this kind of recipients.

Ideally for each case, a target fiber study and transfer and persistence experiments should be carried out using relevant materials to the circumstances. Of course such an approach is practically impossible and the assessment relies rather on the accumulated knowledge derived from experimental studies. For this reason, although the Bayesian approach greatly minimizes the subjective component of the interpretation process, it has never been actually used in the courtroom for fiber evidence. However, the Bayesian approach has been the initiator of numerous experimental studies, and remains a potentially valuable tool for the future.

Conclusions

Background data available to the forensic fiber examiner to assist in interpreting the significance of fiber findings has increased greatly in the past 20–30 years. There is also a deeper understanding of the factors which need to be considered, and acceptance that it is incumbent on the practitioners to assist the court in assessing significance.

A Bayesian approach has been applied in simpler evidence categories, but the complexity of fiber evidence has negated its actual use so far.

Given the circumstances facing forensic fiber examiners in different climatic and socioeconomic environments, local databases and research to put available data in a local context is essential. This type of data is now emerging.

See also: **Evidence**: Statistical Interpretation of Evidence/Bayesian Analysis. **Fibers**: Types; Transfer and Persistence; Identification and Comparison. **Hair**: Comparison: Significance of Hair Evidence.

Further Reading

Biermann TW and Grieve MC (1996) A computerized data base of mail order garments: a contribution toward estimating the frequency of fiber types found on clothing. Part 1: The system and its operation. *Forensic Science International* 77:65–73.

Biermann TW and Grieve MC (1996) A computerized data base of mail order garments: a contribution toward estimating the frequency of fiber types found on clothing. Part 2: The content of the data bank and its statistical evaluation. *Forensic Science International* 77:75–91.

Biermann TW and Grieve MC (1998) A computerized data base of mail order garments: a contribution toward estimating the frequency of fiber types found on clothing. Part 3: The content of the data bank – is it repesentative? *Forensic Science International* **95**:117–131.

Bruschweiler W and Grieve MC (1997) A study on the random distribution of a red acrylic target fiber. *Science and Justice* 37:85–89.

Champod C and Taroni F (1997) Bayesian framework for the evaluation of fiber transfer evidence. *Science and Justice* 37:75–83.

Champod C and Taroni F (1999) Interpretation of fibre evidence. The Bayesian approach. In: Robertson J and Grieve M (eds) *Forensic Examination of Fibres*, pp 379–398. London and Philadelphia: Taylor and Francis

Cook R and Wilson C (1986) The significance of finding extraneous fibers in contact cases. *Forensic Science International* 32:267–273.

Cordiner SJ, Stringer P and Wilson PD (1985) Fiber diameter and the transfer of wool fibers. *Journal of the Forensic Science Society* 25:425–426.

Gaudette BD (1988) The forensic aspects of textile fiber examination. In: Saferstein R (ed.) *Forensic Science Handbook*, vol 2, pp. 209–272. Englewood Cliffs, NJ: Prentice Hall.

Grieve M (1999) Inyerpretation of fibres evidence. In: Robertson J Grieve M (eds) *Forensic Examination of Fibres*, pp 343–363. London and Philadelphia: Taylor and Francis

Grieve MC and Biermann T (1997) The population of coloured textile fibers on outdoor surfaces. *Science and Justice* 37:231–239.

Home JM and Dudley RJ (1980) A summary of data

obtained from a collection of fibers from casework material. *Journal of the Forensic Science Society* **20**:253–261.

Jackson G and Cook R (1986) The significance of fibers found on car seats. *Forensic Science International* **32**:275–281.

Kelly E and Griffin RME (1998) A target fiber study on seats in public houses. *Science and Justice* **38**:39–44.

Kidd C and Robertson J (1982) The transfer of textile fibers during simulated contacts. *Journal of the Forensic Science Society* **22**:301–308.

Palmer R and Chinherende V (1996) A target fiber study using cinema and car seats as recipient items. *Journal of the Forensic Science Society* **41**:802–803.

Parybyk AE and Lokan RJ (1986) A study of the numerical distribution of fibers transferred from blended fabrics. *Journal of the Forensic Science Society* **26**:61–68.

Pounds CA and Smalldon KW (1975) The transfer of fibers between clothing materials during simulated contacts and their persistence during wear. Part I: Fiber transference. *Journal of the Forensic Science Society* **15**:17–27.

Pounds CA and Smalldon KW (1975) The transfer of fibers between clothing materials during simulated contacts and their persistence during wear. Part II: Fiber persistence. *Journal of the Forensic Science Society* **15**:29–37.

Pounds CA and Smalldon KW (1975) The transfer of fibers between clothing materials during simulated contacts and their persistence during wear. Part III: A preliminary investigation of the mechanisms involved. *Journal of the Forensic Science Society* **15**:197–207.

Robertson J, Kidd CBM and Parkinson HMP (1982) The persistence of textile fibers transferred during simulated contacts. *Journal of the Forensic Science Society* **22**:353–360.

Roux C (1997) *La Valeur Indiciale des Fibres Textiles Découvertes sur un Siège de Voiture: Problèmes et Solutions*, PhD thesis. Lausanne: University of Lausanne.

Roux C, Chable J and Margot P (1996) Fiber transfer experiments onto car seats. *Science and Justice* **36**:143–151.

Roux C and Margot P (1997) The population of textile fibers on car seats. *Science and Justice* **37**:25–30.

Salter MT, Cook R and Jackson AR (1987) Differential shedding from blended fabrics. *Forensic Science International* **33**:155–164.

Siegel JA (1997) Evidential value of textile fiber-transfer and persistence of fibers. *Forensic Science Review* **9**:81–96.

Webb-Salter M and Wiggins KG (1999) Interpretation of fibre evidence. Adds to interpretation. In: Robertson J and Grieve H (eds) *Forensic Examination of Fibres*, pp 364–378. London and Philadelphia: Taylor and Francis.

Wiggins KG and Allard JE (1987) The evidential value of fabric car seats and car seat covers. *Journal of the Forensic Science Society* **27**:93–101.

Woltmann A, Deinet W and Adolf FP (1994) Zur Bewertung von Faserspurenbefunden mit Hilfe von Wahrscheinlichkeitsbetrachtungen. *Archiv für Kriminologie* **194**:85–94.

Transfer and Persistence

J Robertson, Forensic Services, Australian Federal Police, Canberra, Australia
C Roux, Department of Chemistry, Materials and Forensic Science, University of Technology, Sydney, Australia

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0514

Introduction

Since the original sin of Adam and Eve, humans have sought to cover themselves with clothing. Hence, the potential for the transfer of fibers has existed from the dawning of humankind. The recognition that it may have forensic value was little recognized until the twentieth century. Although no one individual can actually be credited as the first to single out textiles as a source of trace evidence the much-quoted Locard Exchange Principle is generally accepted as the starting point of the modern era of criminalistics based on the ubiquitous nature of the transfer of trace materials. The Locard Exchange Principle is usually stated as 'every contact leaves a trace'. It follows that if all traces from a crime were available one could reconstruct the steps of the events and follow a trail leading back to an individual or a location. In the real world, even when a transfer has taken place, it may not be detected. There are a number of reasons why this may be the case. The amount of material transferred may be so small it cannot be detected or identified by current techniques. Furthermore, immediately after a transfer the trace material will be subject to loss. The persistence of transferred fibers may be so poor that a transfer cannot be detected at a very short time after transfer. It follows that, in normal circumstances, it is only ever possible to propose that a contact has taken place; it is not possible to prove that a contact has not taken place.

Studies since the 1970s by a number of key groups of workers including Pounds and Smalldon, Grieve, Robertson and Roux, have provided a sound basis for an understanding of the factors to be considered in the interpretation of the transfer and persistence of fibers in forensic investigations.

Transfer

Textile fabrics used in upholstery, carpets and clothing are manufactured using a wide variety of mechanisms and many different types of fibers. The transfer of fibers will be under the influence of the latter factors, fabric construction and fiber type as well as the nature

of the contact. Typically contact will be between individuals wearing items of clothing or between an individual and an item such as a seat or a carpet.

Factors affecting transfer

In the general situation, the following factors have been shown to be important in determining the number of fibers which will transfer during a contact.

Fiber type This is important for both the 'donor' item and the 'recipient' item. Some fabrics can be expected to transfer more fibers than others. For example, fabrics made of wool and acrylic will shed more fibers than fabrics made of polyester fibers. Sometimes the potential for a fabric to transfer fibers is called its 'shedability'. In a case situation it may be useful to assess the shed potential of a donor item. Simple tape lifts may only provide a very rough guide to shed potential. Several authors have proposed the use of simulated contact devices. A rough order for shed potential would be: wool $\geqslant$ acrylic $\geqslant$ cotton $>$ viscose $>$ polyester $>$ nylon.

Shed potential does not depend solely on fiber type. Construction and state of wear of fabrics are also key factors.

Fiber morphology and thickness There is evidence that within one fiber type, finer fibers will transfer in greater numbers than coarser fibers. This probably relates to the greater fragmentation of finer fibers compared to their coarser counterparts. Recent studies have shown that fabrics constructed of microfibers can generate up to seven times more fibers than cotton under the same conditions.

Fabric texture and construction As a general rule for the same fiber type, more fibers will transfer from a coarse than from a smooth fabric. This is, however, a gross simplification. Fabric construction is also important. This involves a wide range of factors. As discussed above, the shed potential is determined by a complex interaction of fiber type, fabric construction and the condition of an item, i.e. how well it wears.

Area of contact As a general rule, the greater the area in contact the more fibers can be expected to transfer.

Number of contacts The number of fibers transferred increases with the number of contacts where the number of contacts is small. With increasing contacts some fibers will transfer back to the donor item.

Force of pressure or contact The number of fibers transferred increases with the force or pressure of contact until a plateau is reached beyond which increased force has no further effect. The force of contact also influences the size of fibers transferred with higher pressure resulting in a greater proportion of short fibers.

Differential shedding

Most studies have shown that with fabrics composed of two or more fiber types they do not necessarily shed fibers proportionate to their representation in the donor fabric. A complicating factor which is sometimes forgotten is that a manufacturer's label may give proportions in terms of weight and not fibre or yarn numbers. The underlying reasons for differential shedding include fabric construction, in which only one fiber/yarn type is on the external, exposed surface of the fabric, and the shed potential of different fiber types.

The need to consider differential shedding has been demonstrated in a number of published case studies. Where the proportion of recovered fibers in a case situation is clearly different from that found (by direct observation) in the putative donor it is incumbent on the forensic scientist to explain this apparent discrepancy. This will usually involve simulation experiments. A further factor complicating the interpretation will be the influence of fiber persistence.

Primary and secondary transfer

In the discussion thus far it has been assumed that the transfer is a primary transfer, that is, a direct transfer from a donor item to a recipient item.

Primary, direct contacts can result in the transfer of hundreds and even thousands of fibers. It is well understood that it is then possible for these transferred fibers to be transferred 'on' during subsequent contacts. A good example would be a person sitting in a cinema seat. The first person leaves behind thousands of fibers on the cinema seat, a second person then sits on the same seat and some of those fibers are transferred on to the clothing worn by the second person. This is a secondary transfer. At least in theory tertiary and subsequent lower order transfers are a possibility. In a case scenario, the forensic scientist must remain alert to the possibility of such transfers. It is often the situation that a suspect will be part of an interconnected group of associates and may have been exposed to the potential for fibers to have arrived on their clothing through a nondirect contact. The interpretation of the location and number of fibers requires caution. A complicating factor is that there is no minimum number of fibers below

which one can identify a secondary or subsequent transfer.

Special cases

Most published studies have been conducted using garments. However, fibers may be transferred from any fibrous surface such as upholstery and carpets. The factors influencing fiber transfer from items such as blankets, bed sheets, seat covers etc. are no different from those outlined for the general case. Transfer from carpets has some different considerations, especially where the recipient surface is a shoe. Carpets are only a subcategory of any other fabric although their construction is perhaps the major factor in determining their shed potential. Shoes are obviously a special case as a recipient surface. The composition and roughness of the sole are important parameters to consider. The mechanism of fiber transfer to a shoe surface is not identical to the mechanism of transfer between clothing fabrics. In some ways fiber transfer to a shoe may be more comparable to transfer to any other physical object. Fiber persistence on shoes is discussed below.

Another special case of fiber transfer is the transfer of fibers from a fabric to the individual wearing the item. The best example of this is from a mask or balaclava often worn in a robbery. Sometimes a robber may discard other clothing. It may be possible through a study of fiber transfer to establish contact between the clothing and a suspect. Fiber recovery from the body of a deceased or an alleged victim and its potential to provide evidence is considered too infrequently.

Mechanism of fiber transfer

There has been considerable theorizing with regard to the underlying mechanisms of fiber transfer but only limited attempts at providing experimental proof. It has been proposed that in the general fabric to fabric situation three mechanisms may be involved:

- transfer of loose fragments already on the surface of the fabric;
- loose fibers being pulled out of the fabric by friction; and
- transfer of fiber fragments produced by the contact itself.

It is accepted that electrostatic attraction of fibers is not an important factor in the general case. However, electrostatic attraction may be a factor in special circumstances. This discussion has focused on the transfer of fiber fragments and not on the transfer of yarns or pieces of fabric torn or removed from a fabric by physical means.

Fiber transfer: a dynamic process

It will often be the case that there will be a time gap between the commission of an offense and the apprehension of a suspect. There is evidence to show that the transfer properties of items can alter with the passage of time. This may be due to wear, washing or other treatments. As a general rule, garments will shed less through time. Caution should be exercised where there is a lengthy time gap between the commission of an offense and suspect items being submitted for examination. This factor also needs to be considered when making a decision as to whether or not to conduct simulation experiments.

Persistence

Persistence is the other half of the equation contributing to whether or not fibers will be found following a transfer. Whatever the number of fibers transferred, and almost irrespective of the nature of the recipient surface, there is an initial rapid loss of fibers. This can be as high as 80% in the first few hours with as little as only a few percent remaining after 24 h (**Fig. 1**). Hence, it is essential that every effort is made to collect relevant items from complainants and suspects as soon as possible after an alleged incident.

In some circumstances fiber loss may not follow the classic loss curve. For example, fibers transferred to car interiors will display the classic curve but the time frame is extended and fibers may be retained for weeks as compared to hours or days with garment to garment contacts. Fibers have also been shown to have been retained for longer periods in open air settings and indefinitely on the bodies or clothing of a homicide victim. In both of the latter examples the weather can play an important role. Other special situations include persistence of fibers transferred to hairs, where grooming, including hair washing is the key factor, and persistence on shoes. In the last example persistence is normally very poor and fibers persist for only minutes unless there are special

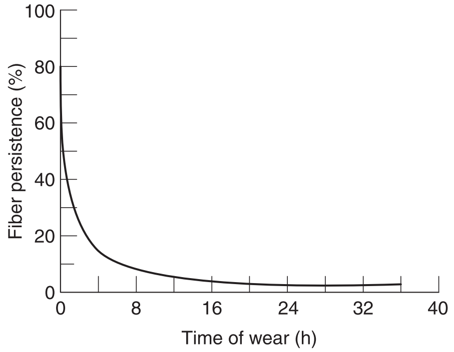


Figure 1 General persistence of fibers on clothing.

reasons such as the presence of a sticky substance or deposit on the shoe sole.

Factors affecting persistence

The loss of transferred fibers will start immediately after the contact which resulted in transfer. A number of factors have been shown to have an influence on the rate of this loss. These include:

- force or pressure of contact – persistence is poorer when the contact is light;
- the location of the contact – fibers are lost more rapidly from areas which are more prone to contact with other surfaces;
- wearing of the recipient garment – fibers are lost more rapidly when the wearer moves after contact; and
- placement of other clothing in contact with area of transfer – fibers are lost more rapidly when other clothing is worn over or on top of the recipient.

Effect of fiber size and morphology

Persistence of short fibers under 2.5 mm in length is greater than longer fibers. A special case is that of microfibers. These show the same overall pattern of fiber loss but have greater persistence compared with 'natural' fibers. This is especially the case with the transfer of microfibers to microfiber garments.

Differential loss

Smooth polyester fibers have been shown to be lost more rapidly than viscose fibers. In general, it cannot be assumed that the rate of loss of different fiber types in a fabric blend will be the same. Hence, over a period of time this factor will add to the potential difference in the ratio of transferred fibers noted earlier. In an extreme example, it may be that one or more type(s) of fiber from a blend may not be recovered. This makes the interpretation of the source fabric difficult unless an explanation can be demonstrated.

Effect of garment cleaning

Studies on the effect of a variety of forms of cleaning have shown that cleaning results in the loss of fibers, but of more importance it is still possible to recover fibers after cleaning. In general, caution needs to be exercised in interpreting the significance of a contact based on the location of recovered fibers because of their potential for redistribution. This is especially the case when a garment may have been cleaned.

Fiber binding and recovery

Three states of binding have been proposed to explain the persistence of fibers after garment to garment contact. These are loosely bound, bound and strongly bound states. It is suggested that loosely bound and bound fibers are lost first and strongly bound fibers become physically trapped in the weave of recipient fabrics. The method of recovery should be selected to maximize the chances of recovering fibers of evidential significance. This choice will depend on the case circumstances.

Conclusions

An understanding of the factors which impact on the transfer and persistence of fibers is critical if the forensic scientist is to interpret what meaning should be attached to the finding of fibers thought to be of evidential significance. As the circumstances of each case are different each case must be considered on its own merits.

Once fibers have been transferred to a particular area of a garment they can also be redistributed over the garment and indeed on to other garments. All of the clothing worn may not have been submitted for examination. If only a small number of fibers is found on items it may be because:

- there has been a long time gap between contact/transfer and examination;
- the fibers have arrived on these garments by redistribution;
- of a secondary or subsequent transfer;
- the recovery method was not efficient; or
- transfer is coincidental and not real.

It follows that:

- because fibers are so readily lost and retransferred, undue significance should not be placed on the exact distribution of a small number of fibers;
- unless a suspect is apprehended fairly quickly, subsequent to an incident, failure to find fibers matching the complainant's clothing does not necessarily imply lack of contact;
- evidence of contact and hence association found through comparison of transferred fibers will generally be of recent ones;
- it is vital to the integrity of fiber evidence that good contamination prevention procedures are in place; and
- as the time of wear increases, those fibers which do remain will be very persistent and difficult to remove, hence efficient methods of recovery need to be used.

The type of information that the forensic scientist should seek should include:

- what is alleged to have taken place – who is involved and how?
- where is the incident said to have taken place? If it

was in a house or in a car, who was the occupier or owner?
- with a sexual assault, did it occur on a bed, on the floor? Is it possible to reconstruct the sequence of events? Were bed covers present and were they moved?
- when did the incident take place and was there any delay before the scene was examined?
- did any person involved have legitimate access to the scene or legitimate contact with the other person or persons before the incident?
- are reliable descriptions available of what was being worn by the offender?
- were items of clothing removed during the incident?

This type of information is necessary if the scientist is to conduct meaningful experiments aimed at reconstructing the events of an alleged incident. There will rarely, if ever, be simple and easy answers to the interpretation of fiber evidence.

See also: **Fibers**: Types; Recovery; Identification and Comparison; Significance. **Hair**: Hair Transfer, Persistence and Recovery.

Further Reading

Kidd CBM and Robertson J (1982) The transfer of textile fibers during simulated contacts. *Journal of Forensic Science Society* **22**:301–308.

Pounds CA and Smalldon KW (1975) The transfer of fibers between clothing materials during simulated contacts and their persistence during wear – part 1: fibre transference. *Journal of Forensic Science Society* **15**:17–27.

Pounds CA and Smalldon KW (1975) The transfer of fibers between clothing materials during simulated contacts and their persistence during wear – part 2: fiber persistence. *Journal of Forensic Science Society* **15**:29–37.

Pounds CA and Smalldon KW (1975) The transfer of fibers between clothing materials during simulated contacts and their persistence during wear – part 3: a preliminary investigation of mechanisms involved. *Journal of Forensic Science Society* **15**:197–207.

Robertson J (ed) (1992) *The Forensic Examination of Fibers*. New York, London: Ellis Horwood.

Robertson J and Grieve MC (eds) (1999) *The Forensic Examination of Fibers*. London: Taylor and Francis.

Robertson J, Kidd CBM and Parkinson HMP (1982) The persistence of textile fibers transferred during simulated contacts. *Journal of Forensic Science Society* **22**:353–360.

Roux C (1997) *La Valeur Indiciale des Fibers Textiles Découvertes sur un Siège de Voiture: Problèmes et Solutions*. PhD thesis, University of Lausanne.

Roux C, Chable J and Margot P (1996) Fibre transfer experiments on to car seats. *Science and Justice* **36**:143–152.

Roux C, Langdon S, Waight D and Robertson J (1998) The transfer and persistence of automotive carpet fibers on shoe soles. *Science and Justice* **39**:239–251.

Salter MT, Cook R and Jackson AR (1984) Differential shedding from blended fabrics. *Forensic Science International* **33**:155–164.

Siegel JA (1997) Evidential value of textile fibre – transfer and persistence of fibers. *Forensic Science Review* **9**:81–96.

Types

C Roux, Department of Chemistry, Materials and Forensic Science University of Technology, Sydney, Australia
J Robertson, Forensic Services, Australian Federal Police, Canberra, ACT, Australia

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0513

Introduction

Fibers are often investigated in forensic science as they are a form of evidence that obeys Locard's Exchange principle, that is, every contact leaves a trace. When dealing with fibers coming from an unknown source (for example, fibers found on a body), it is the examiner's task to first identify these fibers, and then to compare them with a sample of known provinance (for example, fibers from the suspect's garment). Textile fibers occur as a multitude of different types. The forensic examination of fibers requires the scientist to have a thorough knowledge of and experience in the area of fiber technology and classification. The aim of this article is to present and discuss the most common types of fibers likely to be encountered in forensic examinations.

Classification of Fibers

A commonly accepted classification begins by dividing fibers into two groups: natural and manmade (artifical) fibers.

The natural fibers may then be subdivided into three classes: animal (protein), vegetable (cellulose) and mineral (asbestos). Animal fibers are further subdivided into three groups, depending on the protein composition and/or utilization: silk (fibroin), wool (keratin) and hair fibers (also keratin). The

vegetable fibers are also subdivided into three groups, depending on which part of the plant is the source of the fiber: seed fibers, bast (stem) fibers and leaf fibers. The naturally occurring mineral fibers are asbestos.

Artificial fibers are subdivided into three groups: synthetic-polymer fibers, natural-polymer fibers and other fibers. The general classification is shown in **Fig. 1**.

Natural Fibers

Animal fibers

All animal fibers are protein based. The difference between proteins arises from different functional groups pendant to the main chain. For example, silk is composed of the protein fibroin whereas wool and other animal hairs are composed of keratin.

Silk Silk is obtained from a class of insects called *Lepidoptera*. Silk fiber is produced in filament form (>1 km) by the larvae of caterpillars when the cocoons are formed. The principal species cultivated for commercial production of silk is *Bombyx mori*, the mulberry silkworm. Tussah silk is produced by *Antherea pernyi*, the wild silkworm.

In the raw state silk filaments consist of two fibers of fibroin embedded in sericin (silk gum). After commercial de-gumming Bombyx silk fibers are transparent, uniform in width (9–12 μm) with smooth and structureless surfaces (they look like synthetic fibers). Tussah silk fibers are darker in color, coarser and less uniform in width (average 28 μm), with pronounced longitudinal striations.

As a fiber for textile use the outstanding properties of silk are its strength, toughness, high regain, excellent soft handle, resistance to creasing, good draping

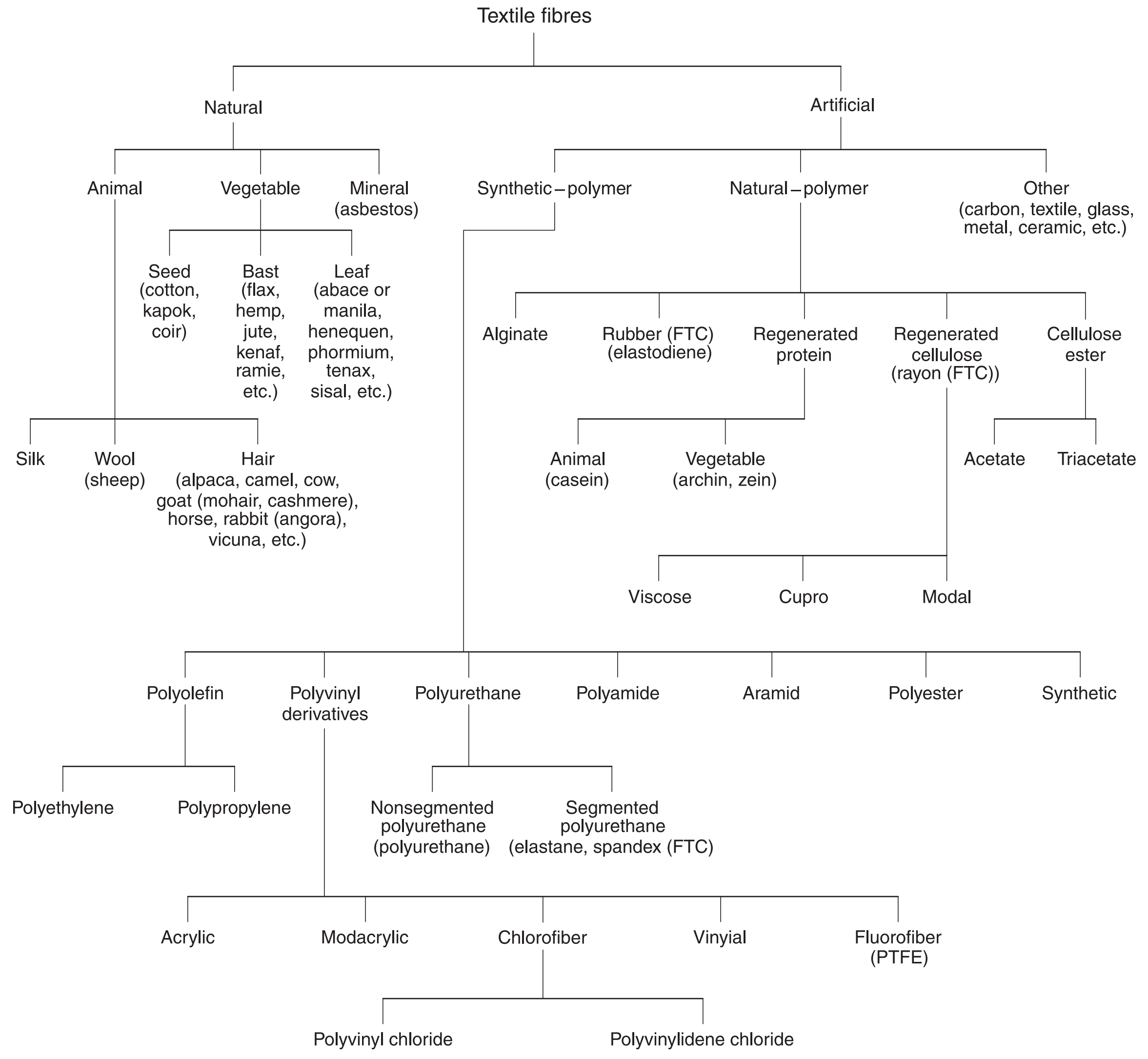


Figure 1 Classification of fibers (courtesy David SK and Pailthorpe MT, 1999).

properties and luxurious appearance. Its high cost has restricted its use mainly to top quality apparel goods such as ladies' frocks, blouses, underwear, stockings, handkerchiefs and men shirts and ties.

An example of silk fiber is shown in **Fig. 2**.

Hair fibers All animal fibers except silk can be considered as hair fibers (in its wider definition). They come from many sources including: ovine (sheep, goat, etc.), camel family (camel, alpaca, vicuna, etc.), or rodent (rat). However, the term 'wool' is generally applied to sheep hair, whereas the simple term 'hair' fibers, or sometimes 'specialty' hair fibers, refers to other animal hairs.

Wool and hair fibers are composed of keratin, which is a protein containing a high concentration of sulfur. These fibers are linear polymers with a high ratio of disulfide bonds which help to stabilize the fiber structure. Hair fibers are resistant to acids and solvents, but are altered by basic and oxidizing solutions.

Hairs are composed of three types of cells: cuticle cells, cortical cells and medulla cells. As shown in **Fig. 3**, the cuticle cells form an outer sheath which encases the inner cortical cells. The cuticle scales overlap each other with the exposed edges pointing toward the tip of the fiber contributing to the tactile properties of the fiber. The number of scale layers and the degree of overlap determines the width of the cuticle. The scale pattern is a useful identification criterion. Scale pattern can vary within different basic hair types (coarse versus fine) within hairs from a single species. Scale pattern can vary in a repeatable and systematic way along the length of individual hairs.

The medulla is usually a centrally located structure, or canal, which can be continuous, discontinuous or fragmented. The structure of the medulla and the ratio between the medulla diameter and the hair diameter (MI = medullar index) are valuable features for the identification of the generic class. It should be noted that, in wool fibers, the medulla is generally present only in coarse fibers (average diameter $>30\,\mu m$).

The most common hair fibers are as follows:

- **Wool:** hair from the sheep. Wool fibers show a great variation in both their physical and chemical properties, due to the multitude of variations possible in diet, breed, health, climate, etc. For example, the length can vary from 3 to 30 cm, and the diameter from 17 to 42 μm (both features are dependent on each other). Merino wool is sought after for its fineness, softness, strength, elasticity and crimp. Merino wool has superior spinning properties and is used for spinning the finest woollen and worsted yarns. Medium wools, produced by breeds such as Leicester, Cheviot, Corriedale and Polwarth are used in the manufacture of woollens, knitting yarns, hosiery, blankets, etc. An example of wool fiber is shown in **Fig. 4**.
- **Mohair:** hair from the angora goat. Mohair fibers have a length of 20–30 cm and a diameter of 10–70 μm. Generally these fibers have no medulla. Mohair has similar physical and chemical properties to wool.
- **Cashmere:** originally hair from the Asiatic goat *Capra hircus laniger*. Hair produced from selectively bred wild goats of Australia, New Zealand

Figure 2 Silk fiber (magnification 90 ×).

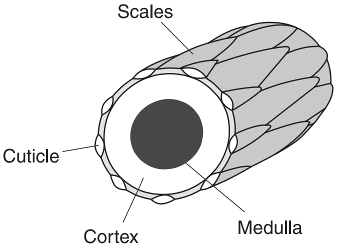


Figure 3 Schematic of a wool/hair fiber.

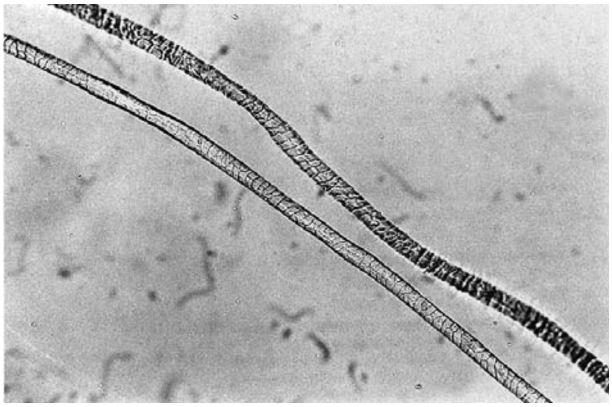


Figure 4 Merino wool fiber (magnification 90 ×).

and Scotland is also called cashmere. These fibers are 5–10 cm in length with diameters of 14–16 μm. Cashmere is chemically identical to wool, but because of its fineness and better wetting properties, is more susceptible to chemical damage, especially with respect to alkalis. Cashmere is higly regarded for producing garments which are comfortable and have soft handle. The main uses are in high class ladies dress goods and knitwear.

- **Camel:** hair from the camel or dromedary. The finer inner camel hair has found use in men's high grade overcoating, whereas the longer outer hair is used mainly in beltings and inner linings.
- **Alpaca:** hair from the fleece of the alpaca (*Lama pacos*) which inhabits South America. Alpaca fibers have a diameter of 24–26 μm, with a distinctive scale structure and medullation. They have similar uses to mohair.
- **Vicuna:** undercoat hair from the vicuna, the rarest and smallest of the llama family. The diameter is 13–14 μm. These fibers are as valuable as cashmere and have similar uses.
- **Angora:** hair from the pelts of the angora rabbit. The hair is shorn from the pelts and separated by blowing. The fine hair is used to make felts for the hat trade, and the long guard hair is spun into yarn. The best hair is 6–8 cm long and about 13 μm in diameter. Rabbit hair is often blended with wool or nylon.
- **Others:** hairs from domestic pets and farm animals (cat, dog, cow, horse, etc.) and hairs from humans (head hairs, pubic hairs, etc.) may be encountered in forensic examinations.

Vegetable fibers

The vegetable fibers are divided into three groups depending on the section of the plant from which they are harvested.

- seed fibers: cotton, kapok and coir;
- bast (stem) fibers: flax, ramie, hemp, kenaf and jute;
- leaf fibers: sisal, henequin, abaca and New Zealand hemp.

The use of the term fiber has different meanings for a botanist and a textile scientist. With vegetable fibers the botanical origin may be a true botanical fiber (bast and leaf fibers) or a seed hair (cotton). Textile vegetable fibers usually contain groups or bundles of single fibers called ultimates. The degree of coarseness of the textile fiber depends on whether or not these groups are separated by physical and/or chemical disruption.

All vegetable fibers contain the same chemical constituents including cellulose and hemicellulose, lignins, pectins, water solubles and fats and waxes. They have, however, distinguishable lengths, bundles and cell wall shapes. In addition, the composition, the general appearance and the surface quality can be modified through the manufacturing process (e.g. mercerized cotton). These fibers have generally a high tensile strength but they are easy to crease.

Cotton Cotton is a fiber attached to the seed of several species of the genus *Gossypium*. Originally from India, cotton is now grown in many subtropical countries. The principal species include *G. barbadense* (South America and Egypt), *G. hirsutum* Central America, USA, 80% of the world production) and *G. arboreum* and *G. herbaceum* (Indian and Arabic countries).

Cotton is composed of approximately 92% cellulose. After maturation, as drying proceeds, the walls of the fiber shrink and collapse, the central hollow lumen becomes smaller and flatter, and the fiber develops convolution (twists). There are from 110 (fine) to 60 (coarse Asian cotton) convolutions per centimeter in cotton fibers. These convolutions improve the flexibility of the fiber and hence the spinning properties of the cotton. The convolutions are an important morphological feature used in microscopic identification of cotton.

The morphological appearance and the crystalline structure can be modified by a special procedure using sodium hydroxide (NaOH) and mechanical elongation. This process, invented by the English chemist John Mercer, in 1844, produces mercerized cotton. Mercerized cotton has greatly reduced convolutions.

Cotton fibers can be dyed with a wide range of dyestuffs including direct, azoic, vat, sulfur, reactive and metal complex dyes. Cotton fibers are resistant to solvents. However, they are damaged by strong acids, oxidizing agents and strong bases.

The mechanical properties of cotton fibers are in the middle range. The use of cotton is, therefore, very wide. As a result, cotton represents more than half of the world's annual quantity of textile fibers. An example of cotton fiber is shown in **Fig. 5**.

Kapok These seed fibers are obtained from the pods of the kapok tree, *Ceiba petrandra* (baobab family). Kapok fibers have an average length of 18 mm and a diameter of 20–30 μm. The fiber is oval in cross-section with a wide lumen and a very thin wall. Kapok fibers are composed of about 65% cellulose with the balance being mainly lignins and hemicellulose. Kapok fibers have a low density and are unsuitable for spinning into yarn. As a result they are principally used as a filling in such products as life-

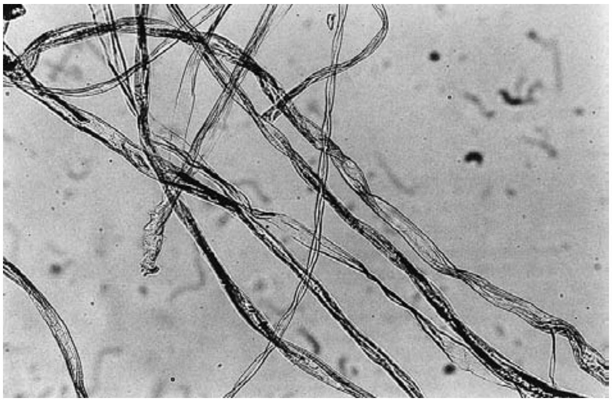


Figure 5 Cotton fiber (magnification 90 ×).

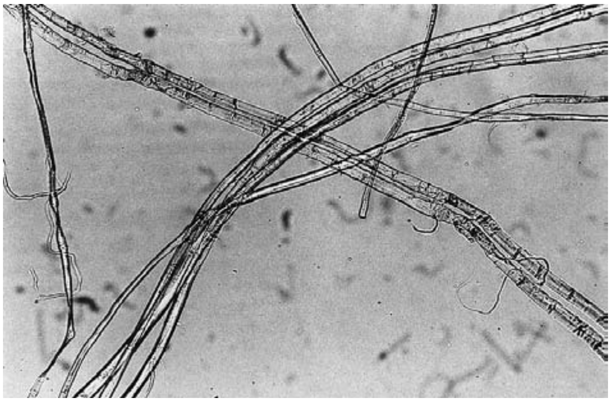


Figure 6 Flax fiber (magnification 90 ×).

buoys, belts, mattresses and pillows. Kapok fibers were popular before World War II (over 40 million kg per annum). However, production has since dropped dramatically owing to destruction of plantations and the development of alternative fiber fillings (e.g. polyester Dacron).

Coir These fibers are contained between the outer husk of the coconut and the shell of the inner kernel. Coir fibers are coarse, with a length of 15–36 cm, and diameters of 0.1–1.5 mm. The virgin coir fibers comprise about 40% cellulose, 40% lignin, 2% pectin and 18% hemicellulose. The process of retting removes much of the pectins and hemicellulose so that the commercial fibers consist essentially of cellulose and lignin. Coir has a high tensile strength which is retained after prolonged immersion in water. As a result, coir is used in the manufacture of brushes, brooms and door mats. Coir also has a good resistance to microbiological attack.

Flax Flax fibers are bast fibers from the stalk of the plant *Linum usitatissimum*. In the flax plant, the fibers are arranged in bundles of 12–40 fibers. These bundles are up to 100 cm in length and 15–20 μm in diameter. Flax fibers are mainly polygonal in cross-section, caused by the way that they are packed together in bundles. Each ultimate fiber is pointed at both ends, and there is a small lumen running lengthwise but disappearing near the ends. The fibers have a smooth surface except at intervals where they are ringed with transverse nodes. These nodes, which are useful in identifying the fiber, help to bind the fibers together, and their regular and frequent occurrence is important for the formation of fine strong yarns. An example of flax fiber is shown in **Fig. 6**.

Flax fibers have a much higher content of noncellulosic material than cotton. Flax contains about 75% cellulose, 15% hemicellulose, 2.5% pectins and 2% lignins. The quality and spinning properties of flax are very dependent on the 1–1.5% wax present. This gives the fibers their high luster as well as imparting suppleness.

Flax fibers are stronger than other natural fibers, and are much less pliable and elastic than other common natural fibers. Flax has about the same moisture regain as cotton but absorbs moisture much more rapidly, making it very suitable for towels and drying cloths. The smooth fiber surface makes it very easy to launder, hence historically flax found great use in the manufacture of table cloths and bandages. Fine fabric woven from flax is termed linen.

Flax is more difficult to dye than cotton. However, flax bleaches to a full white which is enhanced by its natural luster. This characteristic is much prized in household linens, so only a small portion is dyed.

Ramie The commercial source of ramie fiber is *Boehermeria nivea*, which is grown as a perennial. The extraction of ramie is much more difficult than that of flax, jute and hemp fibers, and this difficulty has restricted its use. The ultimate ramie fibers vary in length from 2.5 to 30 cm with diameters ranging from 40 to 75 μm. The cells, which are elongated in cross-section, have thick walls, with a well-defined lumen. The fiber surface is characterized by small node-like ridges and striations. The orientation of the molecules in ramie is very regular, and it is the most crystalline of the natural cellulosic fibers.

Natural ramie consists essentially of about 75% cellulose, 16% hemicellulose, 2% pectins, 1% lignins, and 6% water solubles and waxes. After decortification, the cellulose content rises to about 85%, and the fully de-gummed fiber contains 96–98% cellulose on a dry basis.

Ramie fibers are very white with a silk-like luster and hence make attractive fabrics. Ramie is unaffected by exposure to sunlight and has a very high resistance to bacteria, fungi and mildew. Ramie fabrics and

yarns, like linen, are highly absorbent and dry quickly. Fabrics are easily laundered and show only minor strength loss after repeated washings. Durable and attractive sheets, table cloths and towels can be made from ramie.

Hemp Hemp is the bast fiber produced from the stalk of the plant *Cannabis sativa*. Fiber strands vary from 1.0 to 2.0 m in length with the ultimate cells in the range of 0.5-5.0 cm in length and 15–50 μm in diameter. These ultimate cells are similar in appearance to flax, except that the fiber surface has longitudinal fractures (no nodes) and the cell ends are blunt and irregularly shaped.

The composition of dry hemp fibers is about 75% cellulose, 17% hemicellulose, 1% pectins, 3.6% lignins, 2.7% water solubles and 0.8% fats and waxes. Hemp fibers are gray–green to brown in color with the better quality fibers having a luster. Their strength, elongation, etc. are very similar to flax, and, in some instances, good quality hemp can be used as a substitute for flax. The main commercial uses of hemp are in the manufacture of ropes, cords, string and twine.

Jute Jute is the bast fiber obtained from *Corchorus capsularis* and *C. elitorius* plants. The jute fiber strands comprise bundles of spindle-shaped single cells or ultimates. These ultimates vary in length from 1 to 6 mm, with an average of 3.1 mm, being much shorter than cotton fibers. The diameter is in the range 15–25 μm. They have a polygonal cross-section with a large lumen and thick cell walls. The surface of the cells is mainly smooth with only occasional markings.

The main components of raw dry jute fibers are 71% cellulose, 13% lignin, 13% hemicellulose, 0.2% pectin, 2.3% water solubles and 0.5% fats and waxes. The presence of hemicellulose makes jute more sensitive to alkalis and acids than pure cellulose. Unbleached jute is extremely light sensitive and turns yellow or brown on prolonged exposure with a loss in tensile strength. The sensitivity to light appears to be connected with the high lignin content.

Jute is fairly lustrous with moderate strength. However, it is inextensible and brittle. Jute finds application in packaging for foods, cotton, etc. as backing for linoleum and carpets, and in the manufacture of ropes and cordages.

Kenaf The fiber kenaf is obtained from the plant *Hibiscus cannabinus* and is similar to jute in many of its properties, hence kenaf is a substitute for jute.

Sisal Sisal fibers are obtained from the leaves of the tropical plant *Agave sisalana*. Individual strands of sisal have a length of 1–2 m and are white to yellowish in color. The fiber bundles consist of ultimates which are 3–7 mm in length and have an average diameter of 24 μm. The fibers have a broad lumen and the fiber ends are broad and blunt and sometimes forked. The fibers are polygonal in cross-section, sometimes with rounded edges. The main use of sisal is for the manufacture of ropes, cordages, twines, etc.

Henequin

Henequin is another important agave leaf fiber from the plant *Agave forcroydes*. Henequin is native of the Mexican state of Yucatan. The physical and chemical properties of henequin are almost identical to sisal.

Abaca (Manila hemp) Abaca is produced from the leaf of the plant *Musa textilis*, a member of the banana family indigenous to the Philippines. The fiber strands have a length of 1–3 m and diameter of 0.05–0.30 mm. The ultimate fibers are 3–12 mm in length and 16–32 μm in diameter. The cross-section is irregularly round or oval in shape, and the lumen is very large and generally empty. The fiber cells have thin smooth walls and sharp or pointed ends. Abaca is used for the manufacture of better quality ropes and cordages.

New Zealand hemp (Phormium tenax) *Phormium tenax*, or New Zealand hemp, is a perennial plant indigenous to New Zealand. The ultimate fibers have a length of 2.5–15 mm and diameter of 10–20 μm. The cells are nearly circular in cross-section with a circular lumen. The composition of leaf fibers of *Phormium tenax* differ from other leaf fibers in having much lower cellulose content but higher hemicellulose and lignin contents.

Mineral fibers (asbestos)

The term asbestos applies to a variety of mineral fibers that have a high elasticity and a high resistance to heat, corrosion, wear and tear, humidity, etc. These fibers are noncombustible and have a weak electricity conductivity. Asbestos is easily mixed with building materials (glue, plaster, concrete, etc.). As a result, until relatively recently substantial quantities of asbestos were mined. However, health considerations have since greatly reduced the industry. Asbestos has been replaced wherever possible by other fibers such as glass or Nomex.

The different forms of asbestos vary in composition, dimension and crystalline structure (**Table 1**) However, all forms are silicates. The fibers are very fine (approx. 0.02 μm – 2000 times finer than a human hair) and relatively long (several millimeters).

Table 1 Common forms of asbestos

Group	*Chemical formula*
Serpentine group	
Chrysotile (white)	$Mg_6Si_4O_{10}(OH)_8$
Amphibole group	
Crocidolite (blue, the most hazardous)	$Na_2Fe_3{}^{2+}Fe_2{}^{3+}Si_8O_{22}(OH,F)_2$
Cummingtonite-grunerite (amosite, brown)	$(Mg,Fe^{2+})_7Si_8O_{22}(OH)_2$
Tremolite-actinolite	$Ca_2(Mg,Fe^{2+})_5Si_8O_{22}(OH)_2$
Anthophyllite	$(Mg,Fe^{2+})_7Si_8O_{22}(OH,F)_2$

The different forms can be identified using polarizing microscopy and transmission electron microscopy (TEM).

The most important form of asbestos is chrysotile which accounted for about 90% of total asbestos production. Chrysotile belongs to the serpentine group of minerals and occurs only in serpentine rocks. These are metamorphic rocks, and the development of chrysotile seems to depend on a hydrothermal recrystallization, which is probably initiated at cracks in the rock. The fibers grow at the expense of the adjacent rock. This form is important in a forensic context, because it is used as an ingredient in safe manufacture. When recovered on a suspect's garment, it can therefore constitute a very specific trace evidence.

Artificial Fibers

Artificial fibers may be derived from natural polymers (regenerated fibers such as viscose, rayon, cellulose acetate and triacetate, etc.) or synthesized from simple starting material (synthetic fibers such as polyamides, polyesters, polyurethanes, etc.).

Whether made from a natural or synthetic starting material the aim in making an artificial fiber is to replicate natural fibers but in the process either improve on some limitations of natural fibers or produce an inexpensive alternative. The 'improvements' in artificial fibers often relate to the ability of textile manufacturers to produce fabrics or products with chosen characteristics and the ability to color fibers. The processes used to manufacture artificial fibers are controlled to attempt to produce fibers with defined properties especially with regard to crystallinity and orientation. These contribute to the physical properties of the fibers and the fabric produced from them. The wide range of fibers and fabrics give much greater flexibility and scope to the textile manufacturer than in the era before artificial fibers. However, the fact that cotton still represents more than 50% of textiles and products demonstrates the inability of synthetic fibers to improve on the natural properties of cotton as a textile fiber.

The basis of artificial fiber formation is the same for both regenerated and synthetic fibers. The polymer in a concentrated, viscous form, either in solution or in a molten state, is forced through the tiny holes of a spinneret, and the emerging filaments are immediately precipitated or cooled to form the solid-state fiber. This process is termed spinning or extrusion and may be accomplished in different ways. The fiber filaments thus formed are usually stretched or drawn mechanically. This operation causes the polymer chains to become more aligned (or oriented) in the direction of the longitudinal axis.

The longitudinal and cross-sectional appearance of artificial fibers is largely affected by the manufacturing process. For example, the shape of spinneret holes will determine the cross-section; the orientation of the polymer chains will influence the longitudinal appearance under polarized light, etc.

A guide to most common synthetic fibers is shown in **Table 2**.

Artificial fibers from natural polymers

Fibers can be regenerated from two natural sources:

- a biological source:
  - cellulose: viscose, modal rayon, cuprammonium rayon, cellulose esters;
  - protein: casein;
  - seaweed: alginate fibers;
- a mineral source:
  - glass
  - metallized yarn.

Fibers regenerated from a mineral source are rather classified as 'other artificial fibers' because their source is not a real polymer.

The most common regenerated fibers are cellulosic. Depending on the subclass, these fibers have variable properties (mechanical properties, degree of polymerization, etc.). However, their longitudinal appearance is very similar (uniform diameter with longitudinal striations and irregular cross-section) except cupro (where the surface is smooth and the cross-section is almost round).

Viscose Viscose is the most important regenerated cellulose fiber with respect to its production level. The manufacturing process involves first the aging of cellulose pulp in alkali solution (NaOH). The cellulose pulp is then solubilized by forming the carbon disulfide derivative of cellulose known as cellulose xanthate (cellulose-O-CSSNa). This solution is allowed to

Table 2 Quick guide to artificial fibers (Courtesy American Fiber Manufacturers Association, Washington, DC)

Major generic man-made fibers	*Characteristics*	*Major domestic and industrial uses*
Acetate Celanese, Celebrate, Chromspun, Estron, Microsafe	Luxurious feel and appearance Wide range of colors and lusters Excellent drapability and softness Relatively fast-drying Shrink-, moth-, and mildew-resistant	Apparel: blouses, dresses, and foundation garments, lingerie, linings, shirts, slacks, sportswear Fabrics: brocade, crepe, double knits, faille, knitted jerseys, lace, satin, taffeta, tricot Home furnishings: draperies, upholstery Other: cigarette filters, fiberfil for pillows, quited products
Acrylic Acrilan, Acrilan Plus, Bounce-Back, Creslan, Cystar, Cyster AF, Duraspun	Soft and warm Wool-like Retains shape Resilient Quick-drying Resistant to moths, sunlight, oil and chemicals	Apparel: dresses, infant wear, knitted garments, ski wear, socks, sportswear, sweaters Fabrics: fleece and pile fabrics, face fabrics in bonded fabrics, simulated furs, jerseys Home furnishings: blankets, carpets, draperies, upholstery Other: auto tops, awnings, hand-knitting and craft yarns, industrial and geotextile fabrics
Aramid Kevlar, Nomex, Fi-lana, Pil-Trol, Sayelle, So-Lara, Smart Yarns, Ware-Dated, Wintuk	Does not melt Highly flame-resistant High strength High resistance to stretch Maintains its shape and form at high temperatures	Hot-gas filtration fabrics, protective clothing, military helmets, protective vests, structural composites for boats, sailcloth, tires, ropes and cables, mechanical rubber goods, marine and sporting goods
Lyocell Lenzing Lyocell, Tencel	Soft, strong, absorbent Good dyeability Filrillates during wet processing produce special textures	Dresses, slacks and coats
Modacrylic SEF Plus	Soft Resilient Abrasion- and flame-resistant Quick-drying Resists acids and alkalis Retains shape	Apparel: deep pile coats, trims, linings, simulated fur, wigs and hairpieces Fabrics: fleece fabrics, industrial fabrics, knit-pile fabric backings, non-woven fabrics Home furnishings: awnings, blankets, carpets, flame-resistant draperies and curtains, scatter rugs Other: filters, paint rollers, stuffed toys
Nylon A.C.E., Anso, Antron, Assurance, Avantige, Cantrece, Capima, Caplana, Caprolan,Captiva, Cordura, Crème de Captiva, Crepeset, Durasoft, DyeNAMIX, Eclipse, Hardline, Hydrofil, Hytel, Matinesse, Microsupplex, No Shock, PowerSilk, Resistat, Shimmereen, Silky Touch, Solution, Sportouch, Stainmaster, Stay Gard, Supplex, Tactel, Tru-Ballistic, Ultra Image, Ultra Touch, Ultron, Wear-Dated, Wellon, Wellstrand, WorryFree, Zefsport, Zeftron	Exceptionally strong Supple Abrasion-resistant Lustrous Easy to wash Resists damage from oil and many chemicals Resilient low in moisture absorbency	Apparel: blouses, dresses, foundation garments, hosiery, lingerie and underwear, raincoats, ski and snow apparel, suits, windbreakers Home furnishings: bedspreads, carpets, draperies, curtains, upholstery Other: air hoses, conveyor and seat belts, parachutes, racket strings, ropes and nets, sleeping bags, tarpaulins, tents, thread, tire cord, geotextiles

ripen to an appropriately viscous consistency, and is then extruded through a spinneret. The emerging filaments are immediately coagulated in an aqueous bath containing sulfuric acid and salts of sodium and zinc sulfate. The coagulation bath completely regenerates the xanthate derivative back into the form of cellulose. The filaments thus formed are drawn from the bath at a controlled rate with some degree of applied stretch.

Cellulose is a cellulosic polymer identical to cotton, but with a much lower degree of polymerization. This results in a number of significant differences in the

Table 2 (continued) Quick guide to artificial fibers

Major generic man-made fibers	*Characteristics*	*Major domestic and industrial uses*
Olefin Alpha, Essera, Impressa, Innova, Marvess, Patlon III, Polystrand, Spectra, Synera, Trace	Unique wicking properties that make very comfortable Abrasion-resistant Quick-drying Resistant to deterioration from chemicals, mildew, perspiration, rot, and weather Sensitive to heat Soil resistant Strong; very lightweight Excellent colorfastness	Apparel: pantyhose, underwear, knitted sports shirts, men's half hose, men's knitted sportswear, sweaters Home furnishings: carpet and carpet backing, slipcovers, upholstery Other: dye nets, filter fabrics, laundry and sand-bags, geotextiles, automotive interiors, cordage, doll hair, industrial sewing thread
Polyester A.C.E., Ceylon, Comfort Fiber, Compet, Coolmax, Corebond, Dacronfi, ES.P, Fortel, Hollofi, Kodaire, Kodel, KodOfill, KodOsoff, Pentron, Premafill Plurnp, Premafill Soft, KoSa, KoSa Finesse, KoSa Micronesse, Universe	Strong Resistant to stretching and shrinking Resistant to most chemicals Quick-drying Crisp and resilient when wet or dry Wrinkle- and abrasion-resistant Retains heat-set pleats and creases Easy to wash	Apparel: blouses, shirts, career apparel, children's wear, dresses, half hose, insulated garments, ties, lingerie and underwear, permanent press garments, slacks, suits Home furnishings: carpets, curtains, draperies, sheets and pillow cases Other: fiberfill for various products, fire hose, power belting, ropes and nets, tire cord, sail, V-belts
PBI PBI Logo	Highly flame resistant Outstanding comfort factor combined with thermal and chemical stability properties Will not burn or melt Low shrinkage when exposed to flame	Suitable for high-performance protective apparel such as firemen's turnout coats, astronaut space suits and applications where fire resistance is important
Rayon Beau-Grip, Fibro, Galaxy	Highly absorbent Soft and comfortable Easy to dye Versatile Good drapability	Apparel: blouses, coats, dresses, jackets, lingerie, linings, millinery, rainwear, slacks, sports shirts, sportswear, suits, suits, ties, work clothes Home furnishings: bedspreads, blankets, carpets, curtains, draperies, sheets, slipcovers, tablecloths, upholstery Other: industrial products, medical, surgical products, nonwoven products, tire cord
Spandex Lycra	Can be stretched 500% without breaking Can be stretched repeatedly and returns to original length Light-weight Stronger and more durable than rubber Resistant to body oils	Articles (where stretch is desired): athletic apparel, bathing suits, delicate laces, foundation garments, golfjackets, ski pants, slacks, support and surgical hose
Sulfar Ryton	High-performance fibers with excellent resistance to harsh chemicals and high temperatures Excellent strength retention in adverse environments Flame retardant Nonconductive	Filter fabric for coal-fired boiler bag houses, papermakers' felts, electrical insulation, electrolysis membranes, filter fabrics for liquid and gas filtration, high-performance composites, gaskets and packing

physical and chemical properties between the two polymers. Viscose (as generally with other regenerated cellulosic polymers) is a weaker fiber, with a lower abrasion resistance, lower heat resistance, lower wet strength and poorer heat conductivity than cotton. However, it is the cheapest artificial fiber, and is therefore used for dress fabrics, underwear and furnishings, often blended with cotton, nylon, polyester or wool. Non-woven products include bedding and personal hygiene products, where the moisture absorption is important. Commercial trade names include *Rayonne, Cidena, Rhovalan, Torayon, IF80, Tenasco* (higher tenacity viscose). *Viloft* (hollow viscose). *Darelle* (flame retardant viscose).

An example of viscose fiber is shown in **Fig. 7**.

Modal (polynosic) rayon Modal rayon fibers are obtained by a modified viscose manufacturing process: a very slow regeneration in the presence of dilute

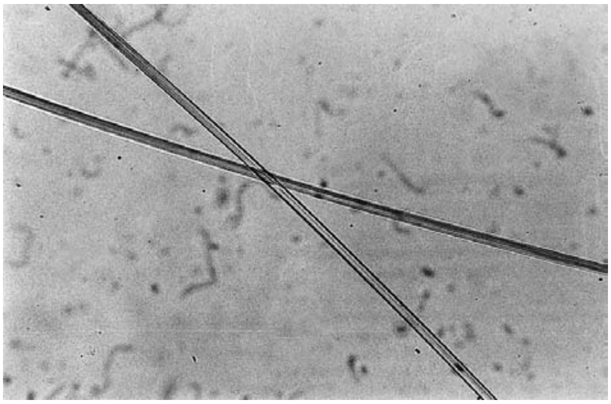


Figure 7 Viscose rayon fiber (magnification 90 ×).

acid and a high concentration of zinc sulfate, accompanied by a very high degree of applied stretch. The highly oriented structure imparts a high degree of strength in both dry and wet states, lower extension at break and lower moisture regain. *Zantrel* is an example of commercial trade name.

Cuprammonium rayon Historically, cuprammonium rayon was the first commercially useful artificial fiber. Cellulose pulp is solubilized in an ammonia solution containing copper salt and caustic soda. The fiber has very similar physical and chemical properties to those of viscose (one minor difference is that the pore size of cupro fibers is larger than that of viscose, so that cupro fibers stain and dye more readily). However, the high cost of cupro production, involving the use of copper salts, resulted in cupro being superseded by viscose in the 1940s. In fact, cupro fiber is now virtually obsolete.

Lyocell Lyocell is a relatively new type of fiber which is defined as a cellulose fiber obtained by an organic solvent spinning process without the formation of a derivative. In this process, raw cellulose is directly dissolved in an amine oxide solvent. The solution is then filtered, and extruded into an aqueous bath of dilute amine oxide, and coagulated into fiber form. Although it is given a separate generic name, the Federal Trade Commission (US) classifies Lyocell as a subcategory under 'Rayon'.

Lyocell fibers are soft, strong, absorbent, wrinkle resistant, biodegradable, and have excellent wet strength. Lyocell produces versatile fabric dyable to vibrant colors, with a variety of effects and textures which simulates silk, suede or leather touch. It is used for dresses, slacks, coats and jeans. Trade names include *Lenzing Lyocell* and *Tencel*.

Lyocell can be distinguished from other cellulosic fibers, including viscose and modal fibers, by a careful examination of their morphology and by accurate birefringence measurements.

Cellulose Ester Fibers (Cellulose Acetate and Cellulose Triacetate) Both acetate and triacetate are ester derivatives of the natural cellulose polymer, in which the hydroxyl groups have been acetylated to form the ester of acetic acid using acetic acid, acetic anhydride and sulfuric acid (as catalyst). The complete conversion of all (6) OH groups in glucose monomer into acetate groups ($-OCOCH_3$) leads to triacetate (or primary cellulose acetate). The partial hydrolysis of the fully acetylated polymer, leaving 74–92% of the hydroxyl groups acetylated, leads to acetate (or secondary cellulose acetate).

In comparison to viscose, cellulose esters are weaker fibers, which become even weaker when wet. Acetate and triacetate materials distort or wrinkle easily. The moisture absorbencies of acetate and triacetate fibers are lower than that of viscose. Both fibers are thermoplastic, that is they may be permanently set or pleated by the application of heat.

Acetate and triacetate fibers are used in most fabric and garment end uses for their silk-like appearance, soft handle and easy dying. Their tenacities are adequate for dress wear, linings and furnishing fabrics, and they are often blended with cotton and synthetics. Trade names for acetate include *Rhodia, Albene, Rhodalha* and *Seraceta*. Trade names for triacetate include *Rhonel* and *Tricel*.

Examples of cellulose acetate fiber are shown in **Figs 8** and **9**.

Regenerated protein fibers Regenerated protein fibers are manufactured by extruding an alkaline solution of protein (casein, peanuts, soya, wool, etc.) followed by coagulation in acid. Further treatment with formaldehyde improves resistance to alkalis. The surface of the fiber is smooth and the cross-section is almost round. Regenerated protein fibers may be distinguished from silk by their insolubility in cold concentrated hydrochloric acid in which silk dissolves. Regenerated protein fibers are usually

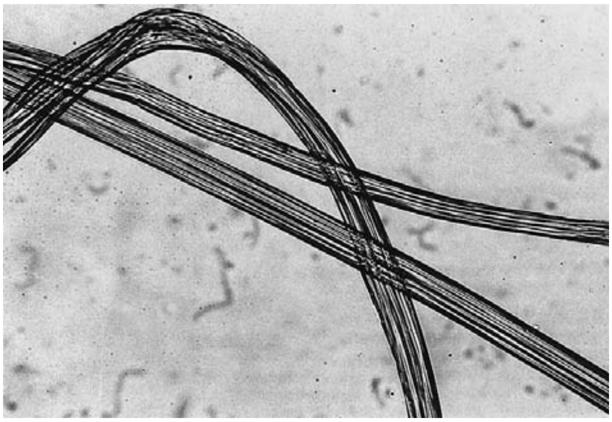


Figure 8 Cellulose acetate fiber (magnification 90 ×).

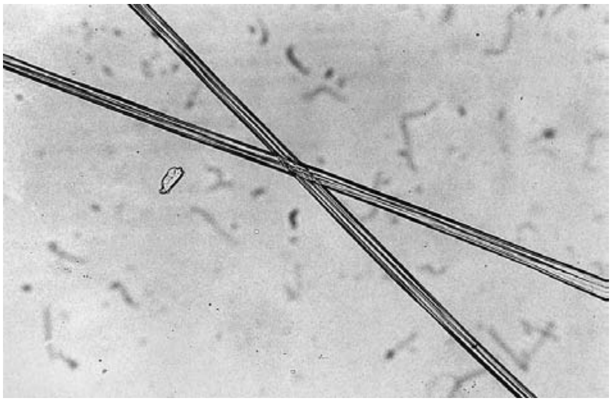


Figure 9 Cellulose triacetate fiber (magnification 90 ×).

incorporated into blends with other fibers such as cotton and wool.

Alginate fibers Alginate fibers are prepared by spinning an aqueous solution of sodium alginate (polymers of mannuronic acid and guluronic acid) extracted from seaweed. The solution is coagulated in a bath of calcium chloride as calcium alginate (precipitate). The fiber is nonflammable owing to its high metal content, and it is soluble in alkaline solution. Alginate fibers may be used in similar end uses as water-soluble polyvinyl alcohol (PVA) (see below). The principal application is, however, in surgical end uses, where the water-soluble form of sodium alginate is readily absorbed in the bloodstream. The appearance of alginate fibers is similar to those of viscose (longitudinal striations, irregular cross-section). Alginate fibers are characterized by their high ash content, and their solubility in carbonate solution.

Synthetic fibers

All synthetic fibers are organic polymeric fibers manufactured by chemical synthesis. The polymer formation is obtained by condensation (loss of water) or addition (no loss), the final product being respectively referred to as condensation and addition polymers. Examples of condensation polymers include polyamide and aramids, polyesters and polyurethanes, and examples of addition polymers include polyolefins and polyvinyl derivatives (cf. below).

The first synthetic fiber, nylon, was developed by Carothers in 1927, but it was only after World War II that mass production started. Nowadays, synthetic fibers are necessary to our daily activity, used alone or blended with natural fibers. They represent an important part of the world production of textile materials, and are *a priori* frequently encountered in criminal investigation. Most common synthetic fibers are described below and a few examples are shown in **Figs 10–12**.

Polyamides (nylon, qiana and aramid) Polyamides are organic polymer fibers which contain an amide functionality (–CONH–) within the repeating unit. The polymer may be aliphatic (nylon), alicyclic (quiana) or aromatic (aramid).

Nylon Aliphatic polyamides are defined according to the number of carbon atoms in the repeating unit. Nylon 6.6 and nylon 6 are the most widely manufactured polyamides (**Fig. 13**).

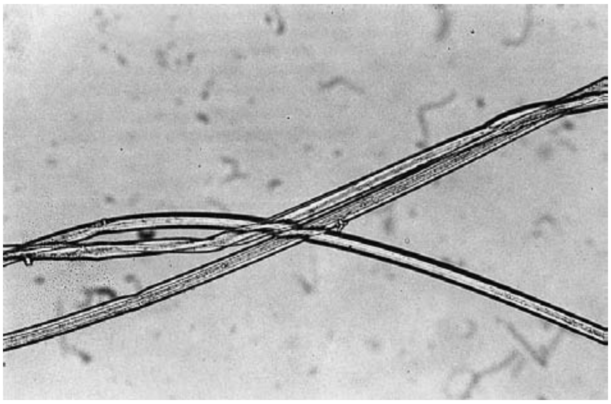


Figure 10 Nylon fiber (magnification 90 ×).

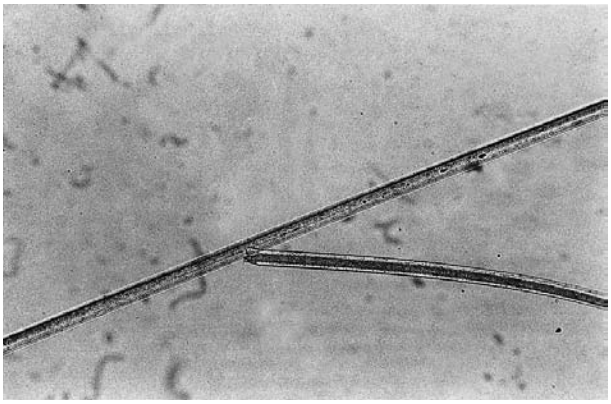


Figure 11 Polyester fiber (magnification 90 ×).

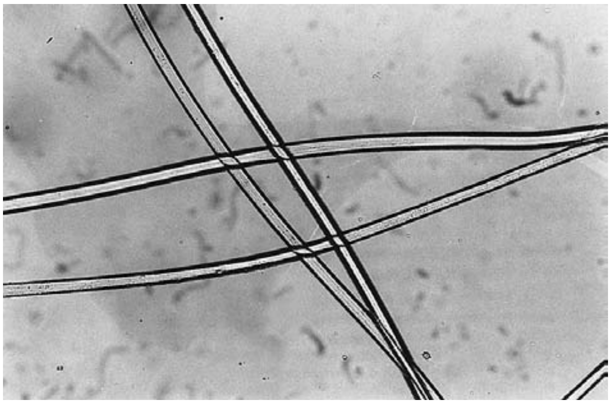


Figure 12 Acrylic (Orlon) fiber (magnification 9 ×).

nylon 6.6

$$HOCO(CH_2)_4CO\left[NH(CH_2)_6NH-CO(CH_2)_4CO\right]_n NH(CH_2)_6NH_2$$

nylon 6

$$NH_2(CH_2)_5CO\left[NH(CH_2)_5CO\right]_n NH(CH_2)_5COOH$$

Figure 13 Two examples of Nylon fibers.

- Nylon 6.6 is the condensation product of hexamethylene diamine and adipic acid. Due to the high degree of crystallinity, nylon 6.6 is the most abrasion resistant of all common textile fibers (four to five times better than wool), with a good tensile strength (even wet) and an excellent elastic recovery. Nylon 6.6 is relatively flammable, chemically stable, but unstable to sunlight. The melting point is in the range of 250–260°C. Nylon 6.6 is commonly used for ropes and cordages, carpets, fishing nets, brushes (e.g. toothbrushes) and common garments (under- and outer garments, often blended with cotton). Trade names of nylon 6.6 include *Blue C, ICI-nylon, Perlon T, Ultron, Nylfrance, Obtel, Oranyl, Amilan*.
- Nylon 6 is the condensation product of 6-aminohexanoic acid. Nylon 6 is easier to dye than nylon 6.6, and is therefore cheaper. The low melting point (approx. 215°C) constitutes a drawback because the garments are difficult to iron. The other properties of nylon 6 are similar to those of nylon 6.6. Nylon 6 is used as tire reinforcement, in carpets, and other similar applications to those of nylon 6.6. Trade names of nylon 6 include *Perlon, Celon, Montefiber, Enkalon, Dederon, Grilon, Lilion, Forlion, Perlofil, Dayan*.

Other aliphatic polyamides that have been commercialized include nylon 4.6, nylon 6.10, nylon 7 and nylon 11 (Rilsan). Some of these fibers have been developed for their exceptional resistance to abrasion and thermal degradation in engineering applications.

New nylon 6-based block copolymer fibers have been developed recently. These fibers have the strength, durability and thermoplasticity of nylon but with a moisture absorption which is similar to that of cotton. Infrared spectrometry is required to distinguish nylon 6-based block copolymers from normal nylon 6.

Quiana Qiana (or PACM-12): is a polyamide containing an alicyclic group and a long 12-membered carbon chain: [–HN–C_6H_8–CH_2–C_6H_8–NHCO–$(CH_2)_{10}$–CO]. This is a fashion fabric known for its silk-like handle and high moisture regain.

Aramids: There are aromatic polyamides defined as having at least 85% of their amide linkage attached directly to two aromatic rings. *Nomex* and *Kevlar* are the most commonly encountered aramids (**Figs 14** and **15**).

Aramids fibers have exceptional mechanical, chemical and heat resistance. *Nomex* is inherently flame retardant and kevlar is five times stronger than steel and more than 10 times stronger than aluminum. However, their expense limits application to high-performance products, for example sporting goods, high performance tires and the aerospace industry.

Polyesters Polyesters are polymers containing at least 85% (mass) of the ester obtained by the condensation of a diol and a dicarboxylic acid. The most common types of polyesters are polyethylene terephthalate (PET, trade names: *Terilene, Dacron, Trevira, Tergal*) (**Fig. 16**) and poly-1,4-cyclohexylene-dimethylene terephthalate (PCDT, trade name: *Kodel*), (**Fig. 17**).

PET fibers have a good tensile strength, and are resistant to chemicals, biological agents and abrasion. They are more flammable than nylon, but more resistant to sunlight. The melting point is around 250°C. PCDT fibers have a good elasticity and are

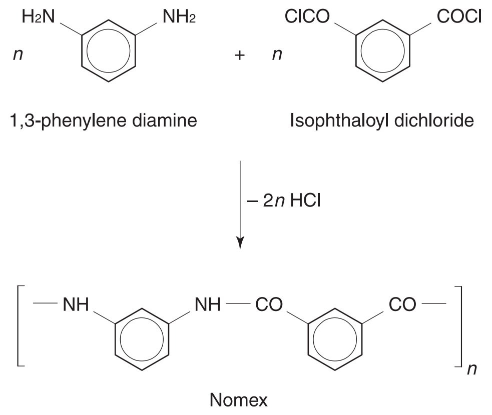


Figure 14 Nomex: benzene rings are *meta*-substituted.

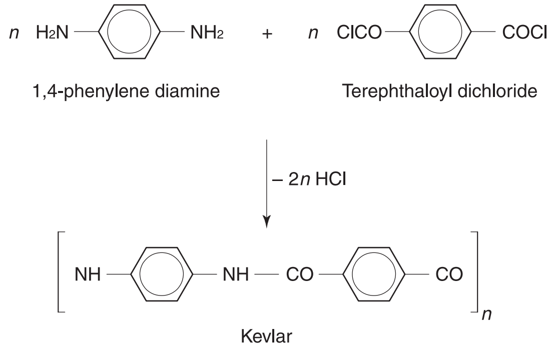


Figure 15 Kevlar: benzene rings are *para*-substituted.

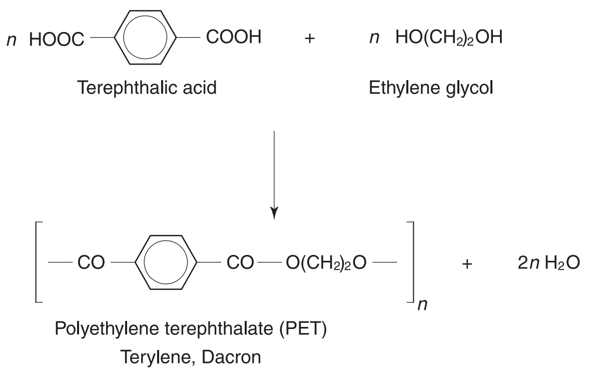


Figure 16 PET fibers.

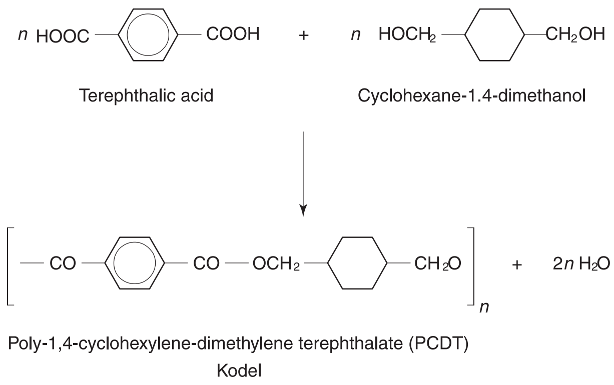


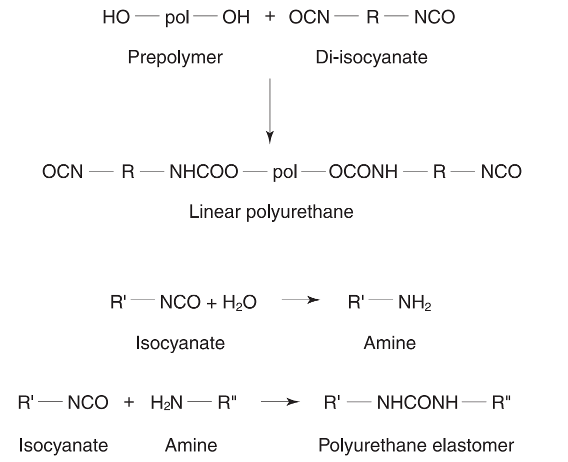
Figure 17 PCDT fibers.

more heat resistant than PET fibers (melting point around 290°C). Their flammability is negligible. Otherwise, their properties are similar to those of PET. Polyesters are used in numerous applications, for example garments (often blended with wool or cotton), linen, ropes, tire reinforcement.

Polyurethanes Polyurethane elastomeric fibers ('elastane') are defined as containing at least 85% by mass of recurrent aliphatic groups joined by urethane linkages [–O–CO–NH–]. Polyurethanes are very complex polymers combining a flexible segment, which provides the high degree of stretch and rapid recovery of an elastomer, with a rigid segment that confers the necessary strength of a fiber (**Fig. 18**).

Polyurethanes fibers have distinctive highly extensible 'snap-back' properties (500–600% elongation). As a result they are often incorporated into fabrics for the purpose of comfort and fit (swimwear, active wear, undergarments, etc.). Trade names include *Lycra, Vyrene, Dorlan, Dorlastan, Lastex S, Rhodastic, Sarlan, Spandelle.*

Polyolefins Polyolefin fibers are defined as containing at least 85% by mass of ethylene, propylene or other olefin units (amorphous olefins excepted). Polyethylene and polypropylene are the two most important types of polyolefins.



Lycra: prepolymer = polyethylene glycol: di-isocyanate = toluyl-2,4-di-isocyanate
Vyrene: prepolymer = polyester: di-isocyanate = diphenyl methane di-isocyanate

Figure 18 Formation of polyurethanes.

Polyethylene Free radical polymerization under conditions of high temperature and pressure, leads to the formation of low-density polyethylene (LDPE) unsuitable as fibrous material. The use of a special catalyst system allows the formation of a highly crystalline polymer network known as high-density polyethylene (HDPE) (**Fig. 19**). Polyethylene fibers have good chemical and biological stability. The melting point is low (around 120°C). They are relatively cheap, light and their flammability is in the medium range (they burn slowly). They have a zero moisture regain. They are sensitive to sunlight degradation. As a result, they have few textile applications. They are used for chemical filtration, car upholstery, medical products and next-to-skin active wear. Trade names include *Courlene, Drylene, Polythene* and *Marlex*.

Polypropylene The manufacture of polypropylene fibers (**Fig. 20**) is much simpler and cheaper than other synthetic fibers such as nylon or polyester. Polypropylene fibers are resistant to solvents and other chemicals, to bacteriological growth and to abrasion. The melting point is relatively low (but higher than that of polyethylene, around 165°C). These fibers are relatively light and have a good tensile strength. They

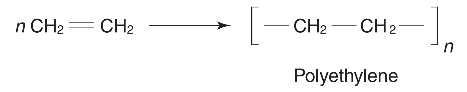


Figure 19 Polyethylene fibers.

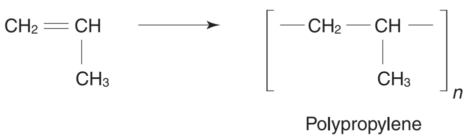


Figure 20 Polypropylene fibers.

have a highly crystalline and hydrophobic nature. They are therefore easily washed and dryed. They constitute an excellent thermal insulation material. The other properties of polypropylene fibers are similar to those of polyethylene. Polypropylene fibers are used as alternatives to natural fibers such as sisal or jute. They can be incorporated in the manufacture of ropes, fishing nets, carpets, furniture fabrics, protective garments, synthetic paper, etc. Trade names include *Aberclare, Deltafil, Fibrite, Gymlene, Herculon, Meraklon, Neofil, Polycrest, Pylen, Reevon, Spunstron* and *Tritor*.

Polyacrylonitrile (acrylics and modacrylics) Polyacrylonitrile fibers are composed of repeating acrylonitrile units [–CH_2CHCN–]. There are two groups of polyacrylonitrile fibers: acrylics and modacrylics.

Acrylics These fibers have at least 85% (mass) of acrylonitrile units (**Fig. 21**). Examples of the most common compositions for acrylic fibers are: acrylonitrile/methyl acrylate, acrylonitrile/vinyl acetate, or acrylonitrile/methyl methacrylate. The purpose of the comonomers is to open the polymeric structure and/or incorporate anionic and cationic groups in the polymer system in order to facilitate the dying process.

Acrylic fibers have a low moisture regain, and a moderate tensile strength compared to nylon and polyester. The resistance to abrasion is moderate (poorer than that of nylon). These fibers are insoluble in most organic solvents. The resistance to poor weather conditions and sunlight is excellent. However, acrylic fibers are very heat sensitive and readily flammable on close approach to a flame. Acrylic fibers are very versatile and have numerous applications, such as outdoors furniture, tents, curtains, carpets, knitted outerwear, usually blended with other fibers to impart flame retardancy. Trade names include *Acrilan, Courtelle, Creslan, Crylor, Dralon, Orlon, Sayelle* and *Zefran*.

Modacrylics These fibers are composed of at least 35%, but no more than 85%, of acrylonitrile units. Common trade names include *Crylor, Dynel, Kanecaron, SEF modacrylic, Teklan, Verel* (**Fig. 22**).

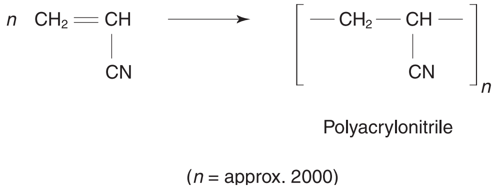


Figure 21 Acrylic fibers.

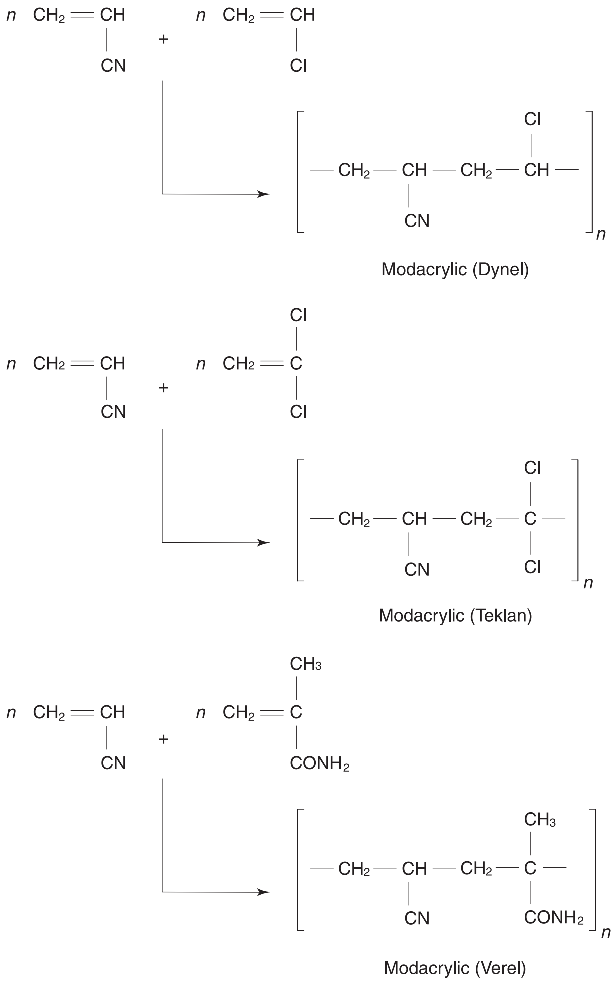


Figure 22 Modacrylic fibers.

Chlorine-containing modacrylics (e.g. *Dynel*) do not support combustion and are considered to be flame resistant, i.e. the flame is extinguished rather than being propagated. However, chlorine-containing modacrylics have lower heat resistance and shrink markedly in boiling water. Modacrylics have similar application to those of acrylics. They are also usually blended with other fibers to impart flame retardancy in apparel and furnishings.

Chlorofibers Chlorofibers are manufactured from either polyvinyl chloride (PVC) or polyvinylidene chloride (PVDC).

Polyvinyl chloride (PVC) These fibers are composed of at least 85% (mass) of vinyl chloride repeating units (**Fig. 23**). PVC fibers have medium tenacity, high elongation, and high degree of shrinkage (up to 40%) at relatively low temperature (100°C). They have virtually zero moisture regain and excellent

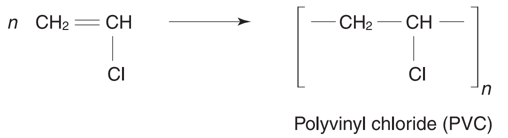


Figure 23 Polyvinyl chloride fibers.

resistance to corrosive chemicals. PVC fibers are also photostable and intrinsically nonflammable. Coloration is however a problem, and they are soluble in chlorinated hydrocarbons and aromatic solvents, rendering them unsuitable for dry cleaning. End use applications are therefore filter fabrics, woven fabrics such as drapery, blankets and underwear, ropes, fishing nets, etc. Trade names include *Clevyl T, Dynel, Kurekalon, Fibravyl, Retractyl, Rhovyl, Thermovyl.*

Polyvinylidene chloride (PVDC) These fibers are composed of at least 80% (mass) of vinylidene chloride as repeating units (**Fig. 24**). PVDC fibers are relatively cheap and their resistance to chemicals is better than that of nylon. They are nonflammable, but shrink at 220°C. They are relatively heavy and extremely water resistant, and have good tenacity. PVDC fibers are used in furniture fabrics, car upholstery (as leather substitute), carpets, etc. The most common trade name for PVDC fibers is *Saran*.

Fluorofibers Fluorofibers are composed of fluorocarbon as repeating units. The most common fluorofiber is polytetrafluoroethylene (PTFE) known under the trade name of *Teflon* (**Fig. 25**). The chain of PTFE forms a highly organized close-packed arrangement. As a result, the fibers have the highest density of any fiber of organic origin. The inertness of the fluorine atoms, combined with the packing symmetry offers a very effective barrier to corrosive chemicals, giving a fiber with outstanding resistance toward heat (stable up to 250°C, nonflammable), chemicals, and solvents. Other properties include a good resistance to abrasion and zero moisture absorbency. PTFE fibers are produced in small quantities for highly specialized end uses, for example filtration of liquids and gases under high temperature and/or corrosive conditions, protective garments. Fluorofibers are also manufactured in Japan under the trade names of *Polyflon* or *Toyoflon*.

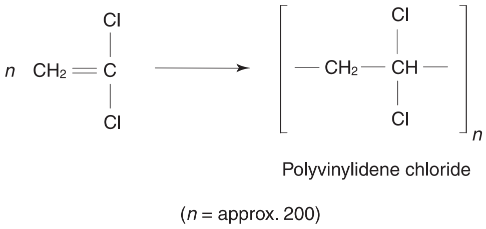


Figure 24 Polyvinylidene chloride fibers.

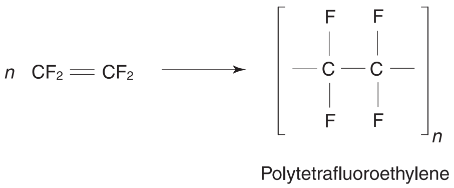


Figure 25 Polytetrafluoroethylene fibers.

Vinylals Vinylals are polyvinyl alcohol chains (PVA) defined as fibers containing at least 50% (mass) of vinyl alcohol and at least 85% (mass) of acetals and vinyl alcohol units. They are produced by the polymerization of the stable vinyl actetate monomer, where subsequent hydrolysis of polyvinyl acetate yields the polyvinyl alcohol polymer (**Fig. 26**). PVA fibers are normally water soluble. However, subsequent acetylation of the OH groups or an after-treatment with formaldehyde renders PVA fibers resistant to boiling water. PVA fibers have generally poor elastic recovery and wrinkle easily. They do not burn easily and have a good chemical and biological stability. PVA fibers exist in a wide range of tenacity. The melting point is around 230°C. PVA fibers are popular in Japan (15% of the total production of fibers) for many textile apparel uses (e.g. kimonos) and industrial applications. They are a good biological resistant substitute for cellulose fibers in fishing nets, ropes, packaging materials, etc. Common trade names include *Cremona, Kuralon, Mewlon and Vinylon.*

Other artificial fibers

Carbon fibers Carbon fibers are prepared from existing artificial fibers, most commonly polyacrylonitrile. Manufacture involves a three-stage heating process during which the fiber is pyrolysed and carbonized leaving a pure carbon fiber. Carbon fibers are

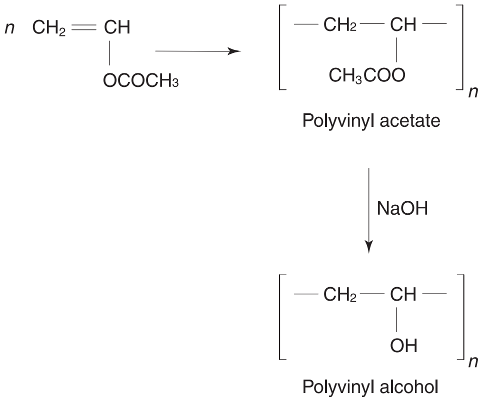


Figure 26 Polyvinyl alcohol fibers.

black and lustrous, and are very resistant to heat, although they tend to oxidize in air at 450°C. The outstanding performance of carbon fibers relates to their stiffness/weight ratio. Carbon complements Kevlar in performance. These fibers are used in the aerospace industry and sporting goods, where stiffness and light weight are essential criteria (overriding the high cost of production).

Glass fibers Glass fibers are manufactured in continuous form by mechanically drawing a molten stream of glass vertically downwards at high speed. Filament diameters are in the range 3–16 μm, and multiple filaments (400–800) are brought together in a strand and processed into yarn, roving, woven, or chopped strand mats. These fibers possess a very high tensile strength, and a good temperature and chemical resistance. They are used for the reinforcement of plastics (boat building, pipes and tanks, etc.) and as filters. Trade names include *Vetrotex, Fiberglas*.

Ceramic fibers Ceramic fibers are a range of inorganic fibers specifically developed for their high temperature resistance. They generally contain either silica or alumina, or proportions of both. Other inorganic fibers also include silicon carbide and boron fibers. These fibers may be spun at very high speed from molten material and are as fine as glass fibers. Fabric applications include hot flue gas filtration, high efficiency air filtration in hospitals, etc., sterilizing filters in the preparation of beverages, thermal and acoustic insulation. Polybenzimidazole (PBI), phenolic (novoloid) and polyphenylene sulfide (PPS) are other inherently flame-retardant fibers which are suitable for protective clothing.

Metallized fibers These fibers are composed of gold, silver, stainless steel, brass containing nickel, aluminum, etc. Polyester films may be coated with a thin layer of vaporized metal (usually aluminum). The opposite is also possible; the two sides of an aluminum foil may be coated with a thermoplastic material. In both cases, the film is subsequently slit into 0.25–1.0 mm wide ribbons. These fibers have a very smooth surface and are usually used for textile decoration (e.g. *Lurex*).

Microfibers Microfibers are extremely fine fibers ($< 7\,\mu m$) developed in Japan in the mid-1970s as substitutes for silk. However they only appeared on the occidental market in the end of the 1980s. Microfibers are most often polyesters, although polyamides, acrylics and viscose microfibers are also manufactured. The textile fabrics containing microfibers are silk-like, lighter and more resistant than traditional fabrics. These fibers are expensive due to the slow manufacture process. Typical applications of microfibers are 'breathing' sport garments which are permeable for water vapour, but constitute an efficient barrier for liquid water. Trade names include *Mitrelle, Setila, Finesse*. It should be noticed that their fine diameter presents a real analytical problem for the forensic fiber examiner.

Bicomponent Fibers Bicomponent fibers are composed of two different components, which can be two different polymers (two different comonomers) or two similar polymers containing different proportions of the same comonomer. The two different components can be assembled in two ways, either side-by-side, or one polymer embedded in the other. Bicomponent fibers are generally acrylics. Trade names include *Acrilan B57* and *Acrilan B94, Cashmilon, Creslan 68, Dralon K, Monsanto X-24, Orlon 21* and *Orlon 78, Velicren*.

Conclusions

Textile fibers occur as a multitude of different possible types. The most common types encountered in a forensic context vary according to different climatic and socioeconomic environments, and also depend on a multitude of factors related to fiber transfer and persistence. However, cotton, wool, regenerated cellulose, polyesters, polyacrylonitrile and polyamide fibers are those most likely to be found in routine cases.

With current analytical techniques the forensic scientist is able to identify the generic class and subclass with a great degree of accuracy. This identification step is crucial in a forensic comparison and requires a great deal of experience and access to thorough reference information (certified samples, relevant atlas, manufacture data, etc.).

The challenge for the forensic scientist in a dynamic environment is to keep the data on the different products, manufacture changes and on how best to discriminate subclasses within a generic type up to date.

See also: **Fibers**: Identification and Comparison; Significance.

Further Reading

Appleyard HM (1978) *Guide to Identification of Animal Fibres*. Leeds: Wool Industries Research Association.

Brunner H and Coman B (1974) *Identification of Mammalian Hair*. Melbourne: Inkata Press.

Carter ME (1971) *Essential Fiber Chemistry*. Fiber Science Series. New York: Marcel Dekker.

Catling DM and Grayson JE (1982) *Identification of Vegetable Fibres*. London: Chapman and Hall.
David SK and Pailthorpe MT (1999) Classification of textile fibers: production, structure, and properties. In: Robertson J and Grieve M (eds) *Forensic Examination of Fibers*, pp. 1–31. New York, London: Taylor and Francis.
Debrot S, Fivaz G, Mermod C and Weber JM (1982) *Atlas des Poils de Mammifères d'Europe*. Institute of Zoology, University of Neuchâtel.
Fairfield CA and Alvey J (1975) *Textile Terms and Definitions*. Manchester: Textile Institute.
Gordon Cook J (1968) *Handbook of Textile Fibres. vol. 1: Natural Fibres*, 4th edn. Watford: Merrow Publishing.
Gordon Cook J (1968) *Handbook of Textile Fibres, vol 2: Man-Made Fibres*, 4th edn. Watford: Merrow Publishing.
Grayson M (1984) *Encyclopedia of Textiles, Fibres and Non Woven Fabrics*. New York: Wiley-Interscience.
Grieve MC (1996) New man-made fibres under the microscope–Lyocell fibres and Nylon 6 block co-polymers. *Science and Justice* **36**:71–80.
Grieve MC, Dunlop J and Kotowski TM (1988) Bicomponent acrylic fibers – their characterization in the forensic science laboratory. *Journal of the Forensic Science Society* **28**:25–33.
Hongu T and Phillips GO (1990) *New Fibers*. Ellis Horwood Series in Polymer Science and Technology. Chichester: Ellis Horwood.
Kroschwitz JI (1990) *Polymers – Fibers and Textiles, a Compendium*. Encyclopedia Reprint Series. New York: Wiley-Interscience.
Lewin M and Preston J (1985) *High Technology Fibers: Handbook of Fiber Science and Technology*. New York: Marcel Dekker.
McCrone WC and Delly JG (1973) *The Particle Atlas*. Ann Arbor: Ann Arbor Science Publishers.
Textile Institute (1975) *Identification of Textile Materials*, 7th edn. Manchester: The Textile Institute.
Warner SB (1995) *Fiber Science*. Englewood Cliffs, NJ: Prentice Hall.

FINGERPRINTS (DACTYLOSCOPY)

Contents

Automated Methods, including Criminal Record Administration
Chemistry of Print Residue
Identification and Classification
Sequential Treatment and Enhancement
Standards of Proof
Visualization

Automated Methods, including Criminal Record Administration

B Dalrymple and **J MacKillican**, Brian Dalrymple and Associates, Orillia, Ontario, Canada

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0522

Introduction

The Automated Fingerprint Identification System (AFIS) is one successful expression of a scientific reality – that no human being is exactly the same as any other human being. Nature does not duplicate. Current science reveals that people (and in a broader sense, all organic entities) differ significantly in almost every way. The ability to reveal and use these differences depends on the discerning capacity of the technology.

First Notice

Scientists and observant students of nature have marveled at the nature and diversity of the skin patterns on their hands since prehistory. Attention

was probably drawn to their fingers for several reasons. Humans have depended on the tool-using ability of the hands in all the primal functions of life, hunting, food gathering, defense and constructing tools and shelter. Also, hands are more easily viewed at close range than other parts of the body. This close scrutiny no doubt revealed the striking contrast between the type of skin found on the fingers and adjacent surfaces (**Fig. 1**).

Over the centuries people have explored the meaning and application of this phenomenon with increasing perception. Images of finger and palm prints adorn the cave walls of early humans. Finger impressions are common on ancient pottery, as artisans completed their craft with a fingerprint signature. In India, business and financial transactions were endorsed with the imprint of the person's finger. In the nineteenth century M. Alphonse Bertillon believed strongly in the individuality of human beings. He attempted to quantify and exploit this phenomenon through a series of detailed and exacting body measurements. The full significance of fingerprint individuality, however, did not fully impact on the forensic world until approximately one hundred years ago.

The progression of fingerprint knowledge and consequently, fingerprint science can be seen in several distinct phases, not necessarily in the following order:

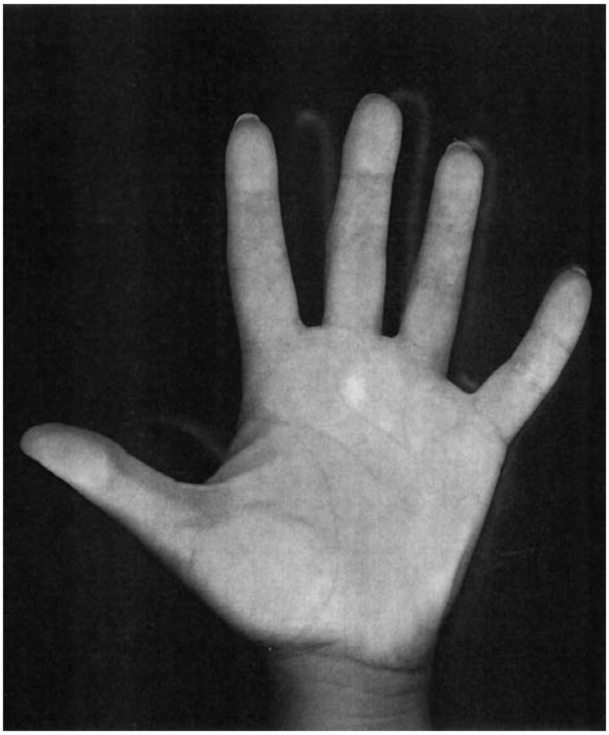


Figure 1 The strange skin on the inside of the hands and fingers.

- Observation and fascination
- Observation of diversity
- Comprehension of personal uniqueness
- Application as a signature on business transactions
- Awareness that the ridge detail does not change over time
- Pattern classification
- Application to criminal investigation – comparison and identification
- Ten print collection and search
- Single fingerprint collection and search
- Palm print collection and search

Ten Print Systems

The first fingerprint classification system was based on the recording of all ten digits. First, the pattern type of each digit was assigned. Secondary traits of each pattern were then recorded. Final overall classification of an individual's fingerprints was dependent on the combination of the primary pattern type, secondary traits and the digit on which they appear.

It must be remembered that recordings of all ten digits were required for classification and addition of a person to the file base. It follows that all ten are required for a search of the file. For the first time in history it was possible to confirm the identity of an unknown person as an individual present in a police file, or positively eliminate him/her, solely on the recording of his/her fingerprints. His/her name (real or assumed), age and physical characteristics were not required for the search. This must be viewed as one of the most significant developments in the infant science of forensic identification.

That said, the value of a ten print system was obviously limited to 'person to database' searches, in which the searcher possessed recordings of all ten fingers. It offered no search option that could link a single fingerprint to either a person, or another single fingerprint.

The Henry system was based on Galton's approach to filing and classification, and was subsequently adopted first by police in England and eventually by most English speaking police forces throughout the world over other systems also developed. It is worth noting, however, that Juan Vucetich of Argentina, also studied Galton's system, as well as a paper authored by Henry de Varigny from the Revue Scientifique. In 1891, he initiated from these sources, a fingerprint system in Buenos Aires, three years before the Galton system was operational in England. This system claimed the first historic criminal fingerprint identification in 1892.

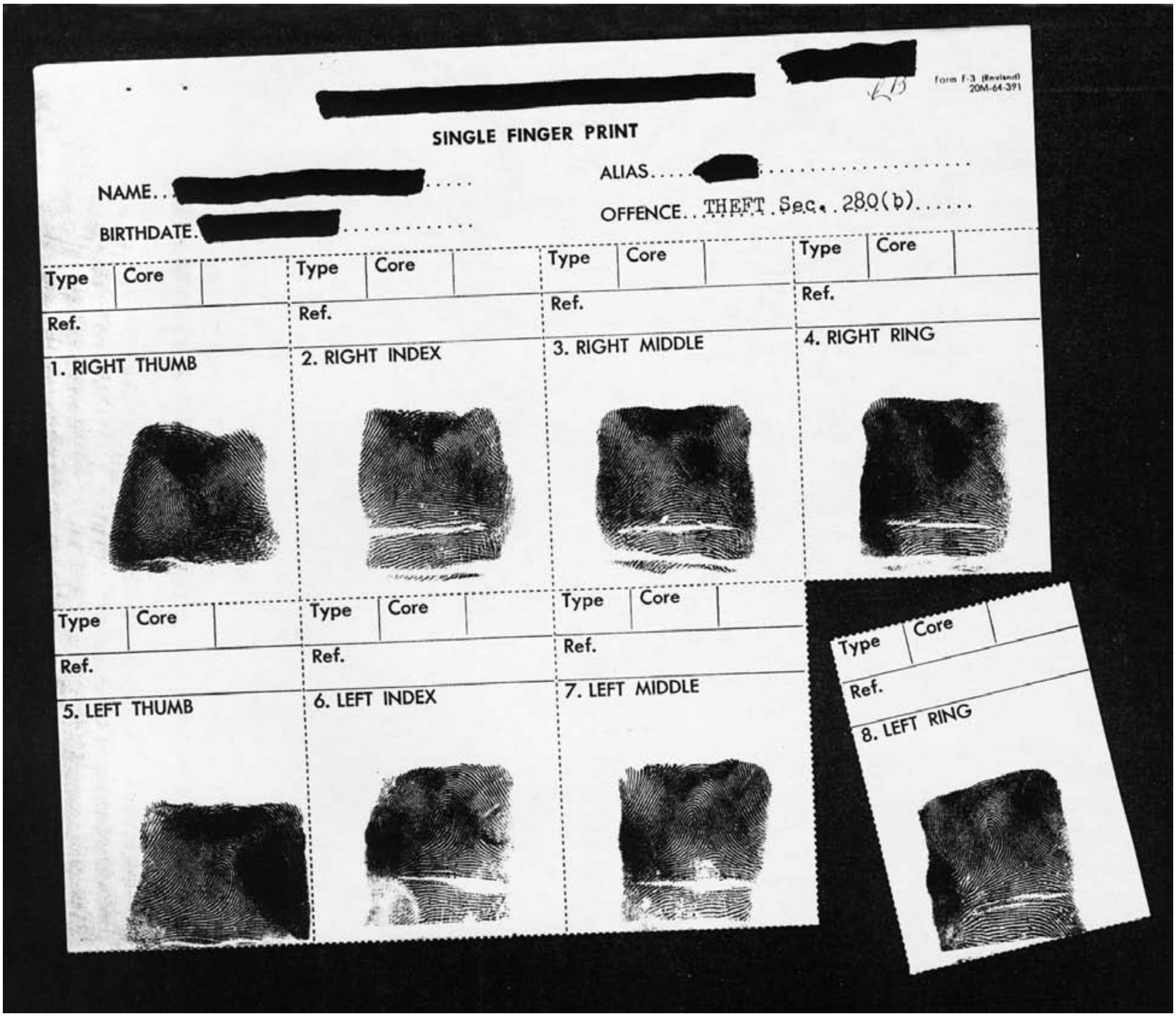
Single Fingerprint Systems

The best known of the manual single fingerprint solutions was the Battley System, which required the following:

- Recording of all digits (except little fingers) on a special, perforated form, which was stiffer than the conventional fingerprint form. The little fingers were omitted from the system because of their relatively rare appearance at crime scenes. This omission streamlined the collection and focused the power of the system on the digits most commonly recovered by evidence technicians;
- Classification of each fingerprint by pattern type and several secondary classifications;
- Separation of the form into the eight component cards, one for each digit. The person's criminal record number was written on each digit card (**Fig 2**);
- Filing of each card in the appropriate section by digit, pattern type and secondary breakdowns;
- Classifying a candidate fingerprint for search by the same system used to create the database;
- When a search resulted in identification, the file number on the card directed the searcher to the complete record information of the individual identified.

Limitations

Computer enthusiasts will be well acquainted with the caveat: 'garbage in, garbage out'. This axiom applies equally well in the area of manual fingerprint search. Users of ten-print systems enjoyed control over recording both the database sets and the set to be searched. This meant that a complete and clear 'nail-to-nail' recording of each finger was possible, assuming that rigid standards of fingerprint recording


[redacted]
Form F-3 (Revised) 20M-64-391

SINGLE FINGER PRINT

NAME..[redacted].....
BIRTHDATE.[redacted]............
ALIAS....[redacted]....................
OFFENCE..THEFT.Sec..280(b)......

Type | Core
Ref.
1. RIGHT THUMB

Type | Core
Ref.
2. RIGHT INDEX

Type | Core
Ref.
3. RIGHT MIDDLE

Type | Core
Ref.
4. RIGHT RING

Type | Core
Ref.
5. LEFT THUMB

Type | Core
Ref.
6. LEFT INDEX

Type | Core
Ref.
7. LEFT MIDDLE

Type | Core
Ref.
8. LEFT RING

Figure 2 Perforated single fingerprint form.

were maintained. Accordingly, if a match existed in the file, the chances were high that a search would find it. In a single fingerprint system, however, the impression for search was developed on an article from a crime scene, deposited unknowingly by a suspect. A clear, complete, 'nail-to-nail' impression is recovered rarely, if ever, from a crime scene or exhibit.

Primary classification based on only the central portion of the print, the part most frequently encountered, was therefore essential. Secondary classification was only useful if the portion of the fingerprint, on which it was based, was present.

- If a critical part of the print were missing, even the apparent pattern type could be different (**Fig. 3**). It can be seen in Fig. 3 that all necessary components required for accurate classification can appear to be present when they are not. In these instances, the technician was required to search all the areas of the collection where the impression might reside. Partial searches, at least, had to be possible when portions of the suspect print were missing. This of course reduced the potential ability of the system to discriminate between different impressions and ultimately, to reveal positive matches.
- The value of any single fingerprint system depends on the elimination at the onset, of the vast majority of the impressions in the database. Unless the search parameters are narrow enough to provide this initial elimination, the number of possibilities requiring comparison will be too large and time-consuming to manage and maintain.

Inconsistency in recording the database impressions presented another problem. Even in controlled ink recording of fingerprints, pressure variations, finger movement and ink thickness routinely contributed to differences in appearance. If these differences resulted in a different classification (as they frequently did), different impressions of the same finger resided in different parts of the database, with an obvious negative effect on the search. It is significant that these problems occurred even within impressions taken under relatively controlled conditions and in the same recording medium – white paper and fingerprint ink. The conditions under which crime scene fingerprints are developed and recorded vary to a far greater extent. The recording medium may include powder, chemically altered sweat, blood or engine oil. The list of surfaces on which fingerprints appear is infinite, but typically includes paper, plastic, glass, metal, wood and skin.

Lastly, the human element must be considered. Even well-trained, experienced fingerprint analysts disagree on classification, when viewing difficult impressions that do not fall obviously into a single category.

All of these factors combined to limit the effectiveness of the manual system and the problem worsened as the database increased in size.

History of AFIS Technology

Computers presented a powerful solution to many of the challenges described above:

- Capability for storing and searching vast quantities of data very rapidly;
- Increased consistency in classification and search criteria;
- Elimination of the need for large volume storage of hard copy fingerprints.

The first AFIS system was conceptualized and developed in 1974 by Printrak, then a division of Rockwell International. The fledgling version of AFIS was

Figure 3 (**A**) Fingerprint clearly revealed as a twinned loop. (**B**) The same impression with a portion missing giving a different apparent pattern type.

adopted by the FBI, with the first ten print card readers installed in 1975. Since then, other large companies have become suppliers of AFIS technology to countries around the world.

Operations

All automated fingerprint systems complete a series of tasks:

- Receive and store 'tombstone' data of each entrant to the system. This includes name, date of birth and other record information;
- Enhance and sharpen ridge detail for maximum accuracy when prints are scanned;
- Automatically scan and classify all ten fingers of each entrant at high speed. (over 100 individuals per hour);
- Compare a suspect impression, classified in the same manner, to the database of known impressions;
- Ignore all the impressions that, on the basis of the data entered, could not be a match and generate a numerical score for each of the top possibilities, based on how closely they resemble the subject of the search;
- Display a match report in descending order, of the top possibilities that cannot be eliminated on the basis of the image data received, and the numerical score assigned to them. In the interests of efficiency, most systems are programmed to display only the top choices;
- One fact must be clearly understood. In spite of the name, all AFIS systems are elimination systems not identification systems. Computers do not, at present, identify fingerprints. The final function of the computer technology in this progression is the presentation of the top scoring possibilities. These possibilities may or may not include a match. The AFIS operator, a highly qualified fingerprint technician, then conducts a comparison using the method in place for a century, and determines the success of the search.

AFIS Search Technology

The approach to classification in automated systems is fundamentally different from that pursued previously in manual single fingerprint collections. They depended on pattern classification, slope, ridge counting and the relationship between core and delta.

The software program used by the computer to perform fingerprint encoding is referred to as an algorithm. When a fingerprint is scanned into an AFIS system, the encoding software allows the computer to 'see' the actual ridge characteristics (bifurcations and ridge endings). The location, or address, of each characteristic is plotted on x–y axes, marked by a circle with a tail to indicate direction (**Fig. 4**). A map is created of the location and direction of all characteristics contained in the scanned impression (**Fig. 5**). The image of the fingerprint is therefore converted into numerical information that is easily stored in the computer memory.

Automated classification does not rely on the pattern designation, slope and core/delta designations that were so problematic in manual collections. It depends entirely on location and direction of minutiae, and their spatial relationship to each other. The advantage of this approach is immediately apparent. The impression in Fig. 3 is a double loop. If this impression appears at a crime scene, the descending loop may not be recorded. The primary classification of the latent would then apparently be a loop, and searching a Battley file would be unsuccessful.

An AFIS search is unaffected by unrecorded loops and deltas, so vital to the effectiveness of a manual

Figure 4 Palm print with the minutiae identified. Courtesy Cambridgeshire Constabulary.

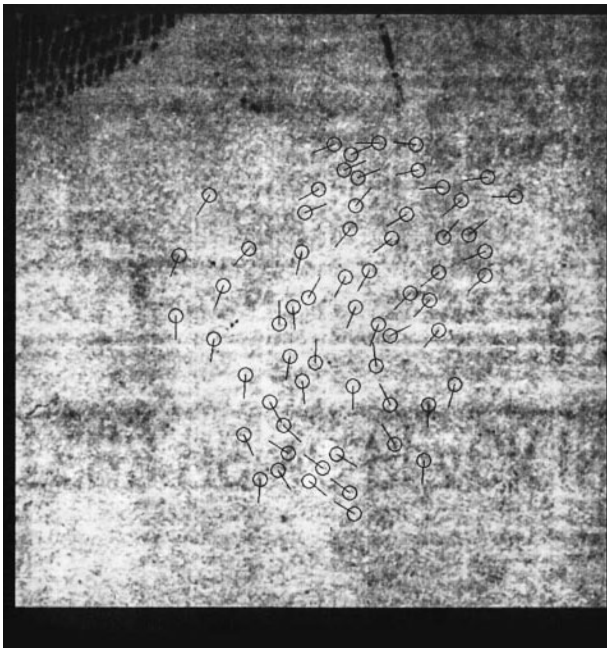


Figure 5 Minutiae map revealing location and alignment of minutiae. Courtesy Cambridgeshire Constabulary.

system. Only the number, clarity and accuracy of the minutiae in the impression impact on success.

Other realities, however, challenge the AFIS search software and consequently, the AFIS user:

- Impressions not placed accurately in the designated space at time of inking arrest fingerprints;
- Uneven or excessive recording medium (e.g. ink);
- The end of the finger is a soft moveable pad. Pressure and twisting during recording can significantly distort the distances between minutiae;
- Permanent and temporary scars;
- All of the above also affect crime scene fingerprints.

As each new version of software becomes increasingly sophisticated and responsive, the impact of these obstacles is reduced. More hits are being made in larger databases in less time.

Search Possibilities

The power and speed of digitization also permit extended searches, some of which were not possible in manual systems.

- **Latent to database:** new latent fingerprints can be searched against the file to establish if that crime can be associated to any subject with a previous record.
- **Latent to latent:** new latent fingerprints can be searched against the file of existing latent impressions to detect connections between crimes.
- **Re-searches:** latent fingerprints associated to serious crimes can be re-searched on a regular basis to determine if the donor, heretofore unknown, has been added to the database in the interim.
- **Regional searches:** searches can be restricted to a specific region when indicated by investigative factors. This reduces pressure on the system operation and shortens the search time.
- **Ten print searches:** a known set of fingerprints is searched against the database to determine if the same person is present under another name.

Negative Searches

If an AFIS search is negative, however, can it be concluded that the person who made that fingerprint is not in the file, and does not have a criminal record? A negative search may be explained by several other possibilities, assuming that an experienced AFIS technician completed the initial encoding of the impression.

- Determination of the digit was wrong and therefore it was not searched correctly.
- Pressure and torque have altered the crime scene fingerprint so drastically as to eliminate the computer's ability to locate a top-scoring match, if one exists in the file.
- The initial recording of the finger was deemed unsuitable for inclusion in the file.
- The finger has been scarred since its initial entry into the file. If this scar has altered the fingerprint sufficiently to lower the score, it will not come to the attention of the operator.
- There may be insufficient area common to both the file impression and the fingerprint being searched. This can result in a numeric score low enough to exclude the print from the top possibilities.
- The fingerprint was searched *after* the subject's arrest, but *before* his fingerprints were included in the AFIS database.

Great care must be exercised to avoid unsupported conclusions.

Livescan

AFIS technicians have historically sought an improvement in ten-print quality. The majority of ten-prints are still recorded manually with ink. As with any manual process, the suitability of prints for digitizing and entry into an AFIS database is dependent on the skill and expertise of the person taking the impressions. Many police agencies are responsible for vast jurisdictions with multiple stations in which fingerprints are recorded. These stations are often far removed from the central location, and from each

other. The submission of these inked ten-prints has relied historically on manual submission, usually by mail or courier, to the AFIS repository. Consequently, there is a delay between the time a subject is recognized and recorded as a criminal, and the time his/her fingerprints become part of the database.

Real time solutions have been developed by the industry in the form of live scan digital ten-print workstations (**Fig. 6**). These devices enable an operator to record fingerprints electronically that, in virtually every instance, are suitable for entry and search within the AFIS database. Livescan stations also check that each fingerprint scanned is suitable for inclusion in the database. The technician is instantly informed if any digits must be re-rolled. To further complete the real time component sought by AFIS specialists, these images are capable of electronic transmission directly to a central AFIS repository. Livescan technology practically eliminates the delay between recording of ten-prints and their addition to the database.

Figure 6 Livescan station for electronic recording of fingerprints. Courtesy Printrak International.

Recognizing that the live scan solution would not be viable for lower volume users, the industry also developed software to allow enabled users to capitalize on simple scanning technology, to digitize and transmit ten print images suitable for import and search in the AFIS repository. Although this solution cannot eliminate the potential inaccuracy of ink recordings, it does facilitate more timely electronic submission to the AFIS database. It must be stressed here that human error in the inking of ten prints can only adversely affect the quality and technical integrity of the database. A poor quality image will not result in an improper identification but could easily result in a successful comparison being missed.

Palm prints

It has been noted that friction skin covers the inner surfaces of both palms and fingers, and is equally valuable for identification in either location. Indeed, palm prints represent approximately 30–35% of the impressions recovered from crime scenes (**Fig. 7**). Use

Figure 7 Latent palm print. Courtesy Cambridgeshire Constabulary.

of palm impressions, however, has traditionally been restricted to comparisons when suspects are known, and when their palm prints are on file.

The taking of palm prints at time of arrest is not a universal practice among police agencies. In 1997, fingerprint experts from the Cambridgeshire Constabulary developed and implemented the first operational AFIS palm print system (**Fig. 8**). This approach has to surmount obstacles even more daunting than those encountered in AFIS fingerprint systems:

- No existing universal protocol for recording of palm prints, similar to the recording of fingerprints, as an automatic consequence of a criminal charge;
- Greater potential for impressions taken poorly or incompletely, and for distortion;
- Much greater surface area than fingers with commensurate increase in the number of ridge characteristics;
- Complete palm prints are rarely found at crime scenes. It is frequently challenging for even experienced fingerprint technicians to determine from which part of the palm the impression originated;
- Correct orientation of crime scene palm prints is much more difficult than for fingerprints. Fingerprint technicians may hold different opinions as to the digit designation of a latent, but rarely will they disagree on the alignment of the impression, or the portion of the finger represented;
- Numerous flexion creases, characteristic of many palm prints, obstruct ridge detail and render it difficult or impossible for the computer to assess satisfactorily.

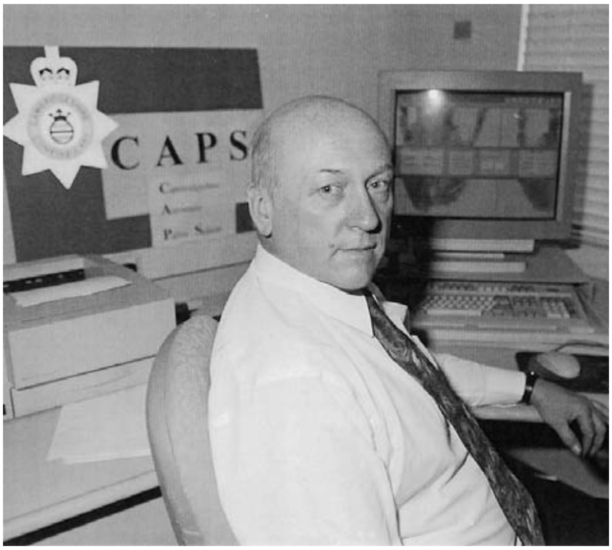


Figure 8 M. Leadbetter, developer of AFIS palm print system. Courtesy Cambridgeshire Constabulary

Administration and Networking

It is vital that record information (including fingerprints) be accessible to investigators at the earliest possible moment. Lawbreakers do not recognize the jurisdictional boundaries of police agencies, and may even exploit them. They have been known to travel hundreds of miles to commit crimes, in an effort to confound investigation. Recent high profile criminals have remained at large and even repeated their crimes, either because their criminal status in another jurisdiction was not known, or it was not known quickly enough in the same jurisdiction.

The final demand of forensic administrators for improved efficiency related to the submission of latent prints directly from the crime scene. Most forensic investigators were hampered by traditional manual lift procedures or at best, were forced to rely on secondary laboratory examination to capture a latent image suitable for transmission to the central AFIS repository for search purposes.

Industry leaders again responded to the demand and developed specialized software that would provide the link for field use. This software provided the bridge that enabled a forensic technician to capture a latent fingerprint using digital photography on location. Once downloaded into a computer hard drive, that image could be transmitted electronically for AFIS search. Forensic investigators can initiate the AFIS search process in real time directly from the crime scene.

Imagine the efficient use this technology offers:

- A technician attends a major crime scene in jurisdiction 'A' and isolates a latent fingerprint.
- This image is captured using digital photography and downloaded to a notebook computer.
- The file is transmitted using the regular or cellular telephone service to a centralized AFIS repository.
- Hours earlier in jurisdiction 'B', a police agency records for the first time the fingerprints of an accused using live scan technology. These digital ten prints are immediately submitted into the same centralized AFIS repository.
- The latent search is conducted and a suspect is positively identified and apprehended, early in the investigation before he can re-offend.
- Both processes happen in real time and have resulted in the forensic investigator providing the information necessary to solve a crime. The efficiencies and effectiveness of this electronic solution in support of forensic identification are significant.

The opportunity to solve crimes quickly allows forensic investigators to effectively manage a greater number of cases. All of these electronic processes

required nothing greater to facilitate transmission than the existing network capabilities accessible to any business today.

In addition, early identification of an offender may eliminate the necessity for hours/days spent tracking leads that, in the absence of forensic evidence, may fail to solve the crime.

There are other benefits on the administrative balance sheet:

- Greater potential to recover property;
- Enhanced disruption of repetitive criminal activity;
- Faster linkages of criminal groups;
- Reduced crime while repeat offenders are in custody.

These components represent significant preventive cost savings and efficiencies to the administrator.

Conclusion

Cost effectiveness and efficient use of resources notwithstanding, one basic axiom should be remembered. Police forces are event-driven, not dollar-driven. The overriding objective of all police activity is the protection of society through timely and effective response. AFIS technology has assisted enormously, and will continue to do so, in the pursuit of that objective (**Fig. 9**).

Figure 9 AFIS technician at workstation. Courtesy of Printrak International.

See also: **Fingerprints (Dactyloscopy)**: Visualization; Sequential Treatment and Enhancement; Identification and Classification; Standards of Proof; Chemistry of Print Residue.

Chemistry of Print Residue

S K Bramble and **J S Brennan**, The Forensic Science Service, London, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0737

Introduction

Fingerprints are still one of the most important contact traces left during the commission of a crime as their uniqueness puts an individual at the scene. Therefore, it is vital that fingerprint recovery bureaus have methods that are effective in the recovery of latent fingerprints. To this end a variety of chemical reagents and associated methods have been successfully developed for visualizing latent fingerprints, some of which are routine and others that are more specialized. Over the years there have been many useful reviews of latent fingerprint recovery techniques and current knowledge of chemical reaction mechanisms and reaction conditions are typically included in these reviews, providing a good starting point for those interested in the science of latent fingerprint recovery. Thus, in order to avoid repetition within the literature this article takes a slightly different slant on the subject by tackling the scientific knowledge from the standpoint of the practical issues involved in the recovery of fingerprints in operational forensic casework. In this way the day to day scientific limitations and difficulties are put into context providing a clear vision of those areas of study that need further investigation.

Before progressing further, it should be noted that fingerprints are also recovered after being deposited in a number of substances, most commonly blood, but these methods are not discussed in this article. In addition, it should be emphasized that the successful identification of an individual from a latent fingerprint relies not only on the quality of recovery but also on the expertise of many disciplines working effectively together. Therefore, as this article concentrates on the chemistry of the process it must be remembered that without high quality photography and pattern recognition the whole recovery process would be fruitless.

This article begins with the current understanding of the chemical composition of natural latent fingerprint residue, followed by the chemistry involved in the recovery of latent fingerprints and finally, the most pressing scientific problems where increased knowledge would be most useful are highlighted.

Chemical Composition of Latent Fingerprint Residue

Intriguingly very little is actually known about the precise chemical constituents of a latent fingerprint at the moment a forensic examiner attempts its recovery. This is due partly to the high number of parameters that affect the final composition but, more importantly, because there has never been any concerted scientific effort towards a greater understanding of the problem.

Traditionally the starting point for a chemical understanding of fingerprint residue composition has been the more rigorously derived composition of those sources deemed responsible for the major compounds present on a fingertip prior to transfer onto a surface namely: sebaceous secretions, eccrine sweat and apocrine sweat. Although, these sources undoubtedly contribute to fingerprint residue they do not provide either a comprehensive list or acknowledge potential chemical activity over the time between deposition and visualization. Since most chemical reagents used in fingerprint recovery procedures are developed on this theoretical model of latent fingerprint residue there is an incentive to pursue a more accurate understanding of the true composition if researchers are to develop new and more effective reagents. Nevertheless, these sources of natural material remain the basis of current knowledge of latent fingerprint residue.

Factors contributing to latent fingerprint residue

Skin surface residue From **Table 1** it might be reasonable to assume that only eccrine sweat and keratins would be found on fingertips, however, through touching the face sebaceous material is also present. These, thus, represent the three main sources of natural material that contribute to the residue found on the surface of a fingertip. Together these natural sources produce a surface film on human skin that is an emulsion of an aqueous and lipid phase (the main emulsifiers being cholesterol and wax esters). Another possible source of material, the apocrine glands, is often included in descriptions of fingerprint residue but is not considered in detail further as their location on the body makes their contribution unlikely or small.

In conclusion there are no definitive data on the chemical composition of a fingerprint and clearly there are some obvious difficulties with ever obtaining such data, not least because every fingertip will have a slightly different surface film mixture. Therefore, the best data available are from the assumption that the film will be a mixture of material from sebaceous glands, eccrine glands and keratinizing epidermis sources as discussed. **Tables 2** and **3** provide a summary of the main chemical constituents in each of these sources.

There have been numerous studies of the individual sources especially sebaceous and eccrine glands. However, care must be taken in interpreting any individual analysis of the glandular material as the means of material collection vary considerably from study to study and some methods discriminate against

Table 1 The main potential contributing sources of natural material found on the fingertips of humans

Source	*Bodily distribution*
Eccrine	All skin surfaces, especially palms and soles
Apocrine	Armpits, chest, abdomen, genitals and soles
Keratinizing epidermis	All skin surfaces
Sebaceous	Forehead, chest, back, abdomen

Table 2 Concentrations of the main constituents of sweat

Compound/electrolyte	*Approximate concentration* ($\mu g\, ml^{-1}$)
Water	–
Lactic acid	2000
Cl^-	1750
Na^+	1350
Amino acids	900
Urea	850
K^+	550
Ammonia	125
Protein	80
Glucose	40
Ca^{2+}	30
Creatinine	10
Mg^{2+}	9
Acetic acid	8
Uric acid	6
Zn^{2+}	1

Table 3 The major lipids present on the skin surface

Lipid	*Surface* (%)
Squalene	10
Sterol esters	2.5
Sterols	1.5
Wax esters	22
Triglycerides	25
Mono- and diglycerides	10
Fatty acids	25
Unidentified	4

certain material. Also, it should be questioned how representative these studies are of the actual material deposited as residue from a fingertip. Furthermore, it is not possible from these data to determine the exact ratio of constituents within a fingerprint. Nevertheless, it is a reasonable assumption that at the moment of deposition some if not all of the compounds in Tables 2 and 3 are present in various amounts. Studies show that a latent fingerprint weighs less than 10 μg and is around 0.1 μm thick.

Deposition The physical contact of the fingertip on a surface is clearly important in determining the initial composition of fingerprint residue. The most obvious parameters involved in this process are time, contact angle and pressure, although it is difficult to determine the precise influence of these parameters in isolation from the nature of the surface itself. Studies have indicated that both the quantity and type of material transferred are dependent on the nature of the surface. Currently, it is not clear whether this transfer process is due to physical or chemical interactions and represents an important scientifically unexplored area of study.

Donor The chemical composition of material found on the fingertips of individuals is known to differ between people but also varies over time with a single individual. There are numerous factors that affect perspiration rates and material secreted. As an example, recent work has indicated that there is a difference in the chemical composition of fingerprints between children and adults. Other factors may include gender, health, diet and medication. Whether these factors have a genuine and important contribution to initial composition has not been investigated scientifically. More excitingly, if gender, age, etc. could be deduced from the chemical analysis of latent fingerprints this could provide important information within the context of a criminal investigation.

As indicated, a major problem in analyzing the chemical composition of a latent fingerprint is in overcoming the problem of sampling. Studies to date usually sample through either washing or wiping the fingertips with a solvent or solvent mixture. Clearly associated information such as donor age, gender, sample area preparation and size of sample group should be documented if meaningful comparisons between studies are to be made. Furthermore, consideration of the solubilities of the various compounds in certain solvents is another important factor. Otherwise there could be difficulties in quantifying results or not collecting certain compounds at all. Only a couple of attempts have been made to sample from a latent fingerprint directly. After sampling a variety of analytical methods have been used to analyze the constituent components including high performance liquid chromatography (HPLC), thin-layer chromatography (TLC), gas chromatography–mass spectrometry (GC-MS), gas chromatography, liquid chromatography and spectroscopy. To date the results are consistent with those expected from the analysis of sweat and sebum. This is not a great surprise as most of the work has relied on samples collected directly from the fingertip. What is needed is to deposit fingerprints and allow them to stand under a variety of conditions commonly encountered in operational forensic casework prior to sampling.

Substrate The substrate onto which the fingerprint is deposited is extremely important for determining the process of visualization. Typically substrates are characterized by two broad types: porous and nonporous. Paper and cardboard typify the former and plastics and metals the latter. Fingerprint residue is usually absorbed into porous surfaces but remains on the surface of nonporous surfaces. These physical (and possibly chemical) effects are still poorly understood. Sweat is more readily absorbed than lipid material therefore the former is more efficiently transferred onto porous surfaces and the latter onto nonporous surfaces. For instance, it has been shown that up to three times more amino acid material is found on absorbent material than similar prints on nonabsorbent surfaces.

Categorization of surface types into porous and nonporous is obviously an oversimplification and in reality there is a range of surface types with varying degrees of porosity. Since the method of enhancement is often highly dependent on surface nature a method for easily and effectively determining physical parameters such as porosity and surface energy would be extremely useful. Such information would allow recovery techniques to be aligned against these substrate physical properties.

Another aspect of surface is its chemical nature. It is reasonable to assume that there may be direct chemical reaction with certain surfaces and also diffusion of chemicals into the fingerprint from the substrate, such as plasticizers. An understanding of this requires further work in the area of surface versus bulk chemical properties. In particular, knowledge on the diffusion depth of different residue components into a substrate is vital in developing reagents targeting surface and/or diffused fingerprint material.

Contaminants Touching involves a two-way transfer of material and everyday activities bring the fingertips into contact with a huge range of materials that could potentially transfer chemical compounds

onto fingertips that, importantly for fingerprint recovery, are trapped in the residue of a latent fingerprint. This is an aspect of fingerprint recovery that has been neglected by scientific study.

Bacteria Skin contains many bacterial species that have largely been ignored by examiners of latent fingerprints, although there have been some attempts to utilize their presence for visualization. It is unknown how important bacteria are in existing recovery mechanisms, in particular their involvement in the subsequent chemistry of the residue after deposition. Inherent fluorescence from latent fingerprints represents a powerful method and the fluorescence mechanism is unknown despite knowledge from other studies that porphyrins produced by bacteria are the origin of the characteristic orange fluorescence of skin.

Ambient conditions and time As indicated earlier there are many potential parameters that could have an effect on the chemical nature of both the substrate and the fingerprint residue after deposition. Time, light, temperature and humidity are some of the more obvious factors but others such as immersion in water or fire damage need to be addressed. This represents an important area for more scientific progress since a more detailed study of these effects would provide a more accurate understanding of the chemicals encountered in latent fingerprints under operational circumstances than is currently available. This is also important in the international arena as climates vary considerably. Since new recovery methods are developed worldwide it is not uncommon for the effectiveness of the methods to vary from country to country presumably due to differences in environmental conditions.

Reagent Chemistry

This section describes the chemical interaction between the latent fingerprint and the methods used for developing a visible print suitable for subsequent pattern analysis.

The photophysics and reagent chemistry of latent fingerprint visualization is largely unknown. Almost all current scientific knowledge is from experiments performed under atypical environments compared to those encountered during the recovery of latent fingerprints. For instance the pathway for the reaction between ninhydrin and amino acids was derived from experiments carried out in solution. There is a requirement for experimental studies on the reaction mechanisms and kinetics of chemical reagents under the typical conditions of latent fingerprint examination, that is, using typical reagents and formulations on real fingerprint residue deposited on both porous and nonporous surfaces. Without this work latent fingerprint recovery will continue to rely on empirical observations and experience in order to optimize reaction conditions. Nevertheless, knowledge from these 'atypical' experiments is vital in providing a lead into this problem.

It is the aim of all latent fingerprint recovery processes to maximize the contrast between the fingerprint and its background. As the variety of methods developed to achieve this goal are too numerous to discuss here, the discussion will concentrate on the factors that govern the success of any chosen method and where current knowledge is lacking. These factors are particularly important when considering the development of new enhancement methods since ideas and techniques developed in the research laboratory need to be developed with an understanding of the practical realities of operational latent fingerprint recovery.

Surface

Since latent fingerprints are inherently invisible it is the surface to be searched that provides the first decision point in choosing a suitable recovery process. Surfaces are traditionally categorized by their perceived porosity based on empirical observation since very little work has been done to establish a method for quickly and reliably characterizing surface properties. It is the properties of the surface that determine the procedures chosen for latent fingerprint recovery as reagents and their formulations have been developed not only for their specificity to certain physical/chemical characteristics of latent fingerprint residue but also for the surface type involved.

As a note of caution the knowledge of the bulk properties of an item may not be indicative of the surface of the item. By way of an example drink cans may be made of aluminum but the surface properties are those of a plastic due to the characteristics of the label printing process.

Ongoing research into surface properties and the development of surface specific reagents is necessary due to the rapid growth of new materials. New surfaces are appearing all the time, such as the development of heat-sensitive papers used in fax machines and tills.

Reagent

As already mentioned there are many chemical reagents used for latent fingerprint recovery (see **Table 4**). In many cases reagents were discovered by accident rather than design. It is most common to

Table 4 Example reagents, formulations, delivery methods and reaction conditions for some common latent fingerprint recovery methods

Reagent/method	*Formulation*	*Delivery method*	*Reaction conditions*
Laser	514 nm and 488 nm	Light guide	–
DFO	Freon:methanol:acetic acid (47:2:1)	Dip	75°C; 15% relative humidity; 30 min
Cyanoacrylate	Cyanoacrylate liquid monomer	Fume	Room temp.; 80% RH; 30 min
Vacuum metal deposition	Gold then zinc	Evaporation	Approx. 10^{-4} Torr
Physical developer	A complex mixture of surfactant, iron and silver salts	Dip	Ambient conditions

categorize reagents in terms of either the surface type or the residue component thought to be the target. Very little is understood about the actual chemical, physical or photophysical mechanisms involved. Strategies for recovering latent fingerprints, therefore, rely on systems derived from practical experience gained over many years. However, it is still important that a deeper understanding of the precise chemical mechanisms involved is obtained for existing methods to be improved. Such knowledge could be vital in answering many observed phenomena, for instance, why do some fingerprints react with ninhydrin and not with DFO (1,8-diazafluoren-9-one), despite the understanding that they both react with primary amino groups? Although it is known that ninhydrin reacts with keratins whether DFO reacts is unknown. Mechanisms for these reactions and others have been described many times and can be found in the literature. A precise understanding of reaction mechanisms would also provide a clear guide to those components within a latent fingerprint that are currently untargeted by recovery methods.

Experience has shown that in order to provide the maximum chance of latent fingerprint recovery from a particular surface more than one reagent is required. This has led to sequences of reagents being set up for the most commonly encountered exhibit surface types. The treatment sequences allow for the recovery of the maximum possible number of latent fingerprints. The various treatments are chosen as they complement each other but they do need to be applied in strict order as certain treatments have an adverse effect on others. Such series of treatments are many and varied but examples of those used within the Forensic Science Service are given in **Table 5**, others can be found elsewhere. In the main, these treatment series have been established empirically.

The chemical reasoning behind such sequences can be best introduced by following one typical sequence of reagents. Table 5 shows that semigloss paper is first examined by light that may either induce fluorescence in the visible and/or ultraviolet. In addition the difference in reflective/absorption properties between the fingerprint and the surface can sometimes be exploited to provide a high contrast image. Cyanoacrylate fuming is applied next as the acetic acid in the DFO and ninhydrin formulations to be used next interferes with the formation of the cyanoacrylate polymer on the fingerprint ridge detail. The precise mechanism of the formation of the polymer from the vapor phase is not fully understood although it is believed that the polymerization is catalyzed by alkaline moieties in the latent fingerprint. Cyanoacrylate fuming is followed by DFO and ninhydrin that react

Table 5 Examples of chemical treatments applied sequentially to surfaces of exhibits at the Forensic Science Service

	Paper	*Semigloss paper*	*Glossy paper*	*Plastic*
Treatment order 1	Light[a]	Light	Light	Light
Treatment order 2	DFO	Cyanoacrylate	Vacuum metal deposition	Vacuum metal deposition
Treatment order 3	Ninhydrin	DFO	Cyanoacrylate	Cyanoacrylate
Treatment order 4	Physical developer	Ninhydrin	DFO	Rhodamine 6G in water
Treatment order 5		Physical developer	Ninhydrin	Crystal violet
Treatment order 6			Modified physical developer	Gentian violet
Treatment order 7				Rhodamine 6G in methanol

[a] Examination by light involves several stages: (1) visual inspection; (2) observation of fluorescence under laser illumination; (3) examination under short- and long-wave ultraviolet light.

with any amino groups present to produce, normally, a fluorescent and purple product, respectively. Finally, physical developer is applied as it is known to react/interact with the lipid fraction although this again is not fully understood.

It can be seen from Table 5 that observation of inherent fluorescence from latent fingerprints is the first treatment in most instances. Although it has been known for some time that fingerprint residue emits fluorescence both in the visible and in the ultraviolet the fluorophore responsible has so far eluded chemical detection. Investigation of possible inherent infrared fluorescence from latent fingerprints has not been attempted.

Reagent formulation

Each reagent has many chemical formulations. These formulations have evolved primarily through empirical observations and experience. Other factors such as cost and to a lesser or greater extent health and safety considerations can also influence formulation. More recently, legislation over chlorofluorohydrocarbons has had a dramatic effect on formulation since Freon (1,1,2,trifluorotrichloroethane) is an important solvent for several reagents currently in use throughout the world. Some examples of reagent formulation are given in Table 4.

A reagent formulation has three important roles. (1) a medium for the reagent itself; (2) to transport the reagent onto/into the surface, and (3) to provide suitable reaction conditions for the reagent. In addition, it is preferable that any formulation does not interfere with subsequent forensic examination of the exhibit. For instance, Freon as a solvent carrier for ninhydrin allows the development of latent fingerprints on paper without causing any handwriting ink to run.

In the case of porous substrates, such as paper, it is necessary to have a formulation that will allow the reagent to penetrate the surface but has a sufficiently rapid evaporation rate to prevent diffusion of the latent fingerprint residue. These issues have to be addressed whilst remembering the health and safety and forensic compatibility constraints mentioned earlier.

Certain methods have been developed to transport the reagent in the vapor phase such as in the cyanoacrylate, DMAC (dimethylaminocinnamaldehyde), fuming and metal deposition techniques.

Reaction conditions

Each reagent and formulation has associated reaction conditions for optimum performance. In some instances these conditions have been optimized for routine work and others for more specific exhibit types. Again, these optimization procedures have been found largely through trial and error and there is a need for more fundamental scientific knowledge on the kinetics of these reactions in order to optimize the reaction conditions.

Table 4 gives some typical reaction conditions for a few of the most common treatments. Between fingerprint recovery bureaus there are subtle variations to these conditions. For instance, most bureaus use cyanoacrylate as a stage in the recovery of latent fingerprints from nonporous exhibits with the majority applying cyanoacrylate in humidified cabinets. Despite its widespread use there has been little work on the flow effects involved in mixing the water vapor and monomer (or oligomers?) in the gas phase. Hence, cabinet design may potentially have a significant effect on the success rate of the method.

It should be noted that the nature of certain casework exhibits, such as vehicles, puts limits on the amount of control that can be placed on the reaction conditions. However, increased understanding through experience has pushed the development of cyanoacrylate fuming tents where vehicles can be fumed under more preferable conditions. Despite these advancements the extremes in ambient temperature determine the relative humidity (hence absolute humidity) that can be achieved which could cause both over- or underfuming.

Forensic compatibility

It must never be overlooked that items that have been treated for fingerprints may be required for examination by other forensic techniques such as DNA or document examination. Therefore, whenever a new reagent is developed its implications for subsequent forensic examination need to be considered. In certain instances reagent formulations have been developed specifically with this in mind, for example Freon as a transport solvent for DFO does not make inks run on paper.

New reagents would also preferably fit into existing treatment series without being detrimental to other techniques in the series. New complementary methods should be able to fit between current treatments or, if they are improved/similar treatments, they could replace existing ones.

Reagent performance

Measuring the relative success of latent fingerprint reagents is one of the most difficult problems facing those developing new reagents, formulations, delivery systems and the application to new surfaces. What is required is a standard against which new methods can be judged. This is not as simple a problem as it

may at first appear. To begin with how do you obtain a 'standard latent fingerprint' that can be used in the tests? Clearly, surfaces could be defined for test comparison purposes but again the properties of deposition of a 'test fingerprint' would need to be overcome. Currently the best indicator of a reagent's effectiveness in the recovery of latent fingerprints is by testing it on a range of 'naturally' handled items such as bank checks or similar objects.

Issues concerning the quality control of reagents and delivery systems also need to be more widely addressed. For instance, some reagents have a shelf-life and its determination is an important aspect of reagent development. Cheap chemical (or other) tests are needed to provide an indication of the quality of a reagent prior to its use to ensure that should no fingerprints be recovered it was not because of a faulty reagent mixture.

The whole issue of quality management therefore has to be addressed; methodologies must be written down and strictly followed, apparatus tested and calibrated regularly, and there should be meaningful quality assurance tests. There are already some systems available but without a more complete understanding of the chemistry/physics some methods may not be operating at their optimum.

Some important unresolved chemical problems in latent fingerprint chemistry

- Recovery of latent fingerprints from skin: to date there is no reliable method for visualization of latent fingerprints from cadavers. Many methods have been attempted but none have led to a routine procedure.
- DNA friendly formulations: as the ability to recover DNA profiles from ever smaller quantities of body fluids or cells increases, the need to ensure compatibility of the recovery method with subsequent DNA techniques is vital. This may not only require the back testing of current systems but could necessitate the development of new reagents/formulations.
- Testing regime for novel reagents/methods: new reagents/methods for latent fingerprint detection are being published constantly but there is no systematic method for their testing. Ideally any test would provide information on sensitivity, residue specificity, and compatibility relative to the best existing techniques.
- Substrate characterization: a simple, robust, reliable and cheap method for characterizing surface properties such as porosity would be a valuable tool in deciding treatment sequences to be applied.
- Fingerprint residue composition: a thorough understanding of latent fingerprint chemical composition would be desirable in order to determine new target compounds for reagents and for optimization of current techniques.
- Aging fingerprints: a method allowing determination of the age of a latent fingerprint could provide vital information in the course of an investigation.
- Reaction mechanisms and kinetics of reagents with latent fingerprints: knowledge of reaction mechanisms and kinetics of the reactions of current reagents with latent fingerprints under typical operational conditions would allow for the optimization of techniques.
- Quality assurance tests: realistic methods for testing the quality of reagents and systems for latent fingerprint recovery would provide confidence and maintain quality between bureaus.
- Event order: methods for establishing the order of events, such as whether the fingerprint was present before/after the handwriting or which of the superimposed fingerprints was deposited first would be of forensic value.

Conclusion

Fingerprints are still and will continue to be a vital evidence type available to law enforcement agencies and therefore the need for continuing research is essential in order to improve on existing methods and to meet the challenges of future needs.

Hopefully, it has been made clear that the recovery of latent fingerprints is not a straightforward chemistry problem but is a complex scenario requiring the consideration of many interrelated factors. This provides a rich and interesting scientific field of endeavor with many questions still to be answered. It was the aim of this article to establish where these questions lie currently and to stimulate activity to provide the answers.

See also: **Crime-scene Investigation and Examination**: Fingerprints. **Fingerprints (Dactyloscopy)**: Visualization; Sequential Treatment and Enhancement. **History**: Fingerprint Sciences.

Further Reading

Allman DS and Pounds CA (1991) Detection of fingerprints on skin. *Forensic Science Review* 3:83–89.

Bobev K (1995) Fingerprints and factors affecting their condition. *Journal of Forensic Identification* 45:176–183.

Burke J (1996) *New Reagents for the Detection of Latent Fingerprints*. PhD Thesis, University of Leeds. Leeds, UK.

Creer KE and Brennan JS (1987) The work of the serious crimes unit. *Proceedings of the International Forensic Symposium on Latent Prints*, p. 96. Quantico, VA: FBI Academy.

FBI (1994) *Chemical Formulas and Processing Guide for Developing Latent Prints*. Washington, DC: Federal Bureau of Investigation, Laboratory Division, Latent Fingerprint Section.

Goode GC and Morris JR (1983) *Latent Fingerprints: A Review of Their Origin, Composition and Methods for Detection*. Report No. 022/83. Aldermaston, Berkshire, UK: Atomic Weapons Research Establishment.

Kent T (ed) (1998) *Manual of Fingerprint Development Techniques: A Guide to the Selection and Use of Processes for the Development of Latent Fingerprints*, 2nd edn. Sandridge, UK: Police Scientific Development Branch.

Lee HC and Gaensslen RE (eds) (1991) *Advances in Fingerprint Technology*. New York: Elsevier Science.

Margot P and Lennard C (1994) *Fingerprint Detection Techniques*, 6th edn. Lausanne: University of Lausanne.

Menzel ER (1989) Pretreatment of latent prints for laser development. *Forensic Science Review* **1**:43–62.

Pounds CA (1986) Developments in fingerprint visualisation. In: Maehly A and Williams RL (eds) *Forensic Science Progress*, vol 3. Berlin: Springer-Verlag.

Rothman S (1954) *Physiology and Biochemistry of the Skin*. Chicago: University of Chicago Press.

Thomas GL (1978) The physics of fingerprints and their detection. *Journal of Physics E. Scientific Instrumentation* **11**:722–731.

Identification and Classification

B Dalrymple, Brian Dalrymple and Associates, Orillia, Ontario, Canada

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0520

Introduction

Humans have been aware of and fascinated by the patterns on their fingers for thousands of years, although there is no conclusive evidence of the first time that the unique nature of these impressions was understood. Children today who finger paint exercise a similar curiosity about the strange and beautiful patterns of lines they have created. The evolution of fingerprint classification and identification as a science has supplied not only comprehensive knowledge of what fingerprints are and how they are formed, but also their full potential as a forensic tool.

History

The following timeline illustrates the progression from fascination to friction skin classification and identification as it is practiced today.

- 1955–1913 BCE: Finger seals on contracts are attributed to the reign of Hammurabi in Babylon.
- Second or third century BCE: Clay seals bearing fingerprints are used by the Chinese.
- Twelfth century: A Chinese novelist makes reference to the use of fingerprints in criminal identification.
- 1684: In England, Nehemiah Grew described the pores and their functions in the skin of the hands and feet.
- 1686: Marcello Malpighi, an Italian university professor, conducted the first friction ridge research under the newly invented microscope.
- 1788: An unsubstantiated report that a German doctor, J.C.A. Mayer, stated his belief in the unique nature of fingerprints.
- 1823: Johannes Purkinje, a German professor, classified the patterns of the finger impressions but ascribed no identification value to them.
- 1860s: In India, Sir William Herschel initiated the taking of fingerprints in efforts to control fraud in pension payments. He also established the immutability of ridge detail by taking his own fingerprints over a long period of time.
- 1880: Dr Henry Faulds concluded that the ridge detail of the fingers was unique. He further claimed to have established identity through comparison of latent fingerprints to known samples.
- 1892: Sir Francis Galton published *Fingerprints*, a book in which he supported Herschel's findings regarding the unchanging, and hence reliable, nature of ridge detail.
- 1891: Juan Vucetich developed and implemented a workable fingerprint identification system in Argentina.
- 1892: In Argentina, the first murder case in the world to be solved by fingerprints.
- 1896: Argentina initiated the first criminal record registry based on fingerprints.
- 1896: Sir Edward Henry created the Henry system of ten print classification, which has persisted as the ten print classification system of choice into the middle of the twentieth century, until the advent of computerized filing systems.
- 1901: First Henry Fingerprint Classification System at Scotland Yard.
- 1912: Edmond Locard, Lyons, France, credited with poroscopy, the science of identification through the position and structure of pores.

- 1930s: Single Fingerprint System developed by Battley, New Scotland Yard

Anatomy

Dactyloscopy (from the Greek word for finger) is the study of fingerprints. Specialized ridged skin, called friction skin, is found exclusively on the palmar surfaces of the hands and the plantar (sole) surfaces of the feet. These are referred to collectively as volar surfaces. Friction skin is designed for grasping and holding objects. All primates have friction skin on the palmar and plantar surfaces.

Ridgeology, a term which has recently entered the lexicon, is more comprehensive in the description of friction skin science because it is not restricted to the tips of the fingers. Although finger and palm prints are the focus of this article, all information concerning identification applies equally to the friction skin of the feet. Friction skin differs significantly in structure and function from the skin covering the rest of the body as follows.

1. It is hairless.
2. It contains no sebaceous (oil) glands.
3. It has a much higher concentration of nerve endings.
4. It has a much higher concentration of sweat glands.
5. There is a lack of pigmentation.

Structure of Friction Skin (Fig. 1)

Epidermis This is the outer layer of tough ridged skin bearing only sweat pores; no vessels, nerve endings or sweat glands. It functions as an aid in gripping and as a protective cover for the delicate components of the dermis.

Dermis This contains the sweat glands, nerve endings and the papillary layer which determines the unique fingerprint code.

Papillae These are pegs or protuberances, arranged in twinned rows on the surface of the dermis. Each double row of papillae on the dermis is represented by one ridge on the epidermis.

Stratum mucosum This is the lower level of the epidermis in contact with the papillary layer.

Eccrine glands More commonly referred to as the sweat glands, the eccrine glands excrete perspiration, thus accomplishing two principal objectives: expulsion of waste material through the pores; temperature regulation, through the cooling effect of perspiration. On the friction skin of the hands an additional purpose is served. The resulting moisture on the friction ridges renders the fingers more efficient in gripping, manipulating and separating objects. Licking one's finger to turn a page utilizes the same principle. Although approximately 99% of the excretion is water, the remainder is composed of various organic salts, nitrogenous compounds and amino acids. These ingredients are vital to the process of detecting fingerprints. When the friction skin of the hands comes in contact with an object, some of the material present on the skin is transferred, placing an invisible image of the detail on that surface, in the manner of a stamp pad. This analogy is even more appropriate if visible material such as ink, paint or dirt is transferred from the fingers. In fact, the taking of fingerprints with ink and paper is merely the controlled and optimized application of actions we take every day when we inadvertently leave our fingerprints on objects we touch. The amount and composition of finger deposit varies greatly as a result of personal metabolism, climate, diet and mood. It is therefore not surprising that some people are more reliable donors than others.

Classification

Elements of a fingerprint (Fig. 2)

- **Recurve:** A recurve is any ridge retracing its course, and includes the bend in a loop or the circular course of a whorl. Arches and tented arches lack free recurves.

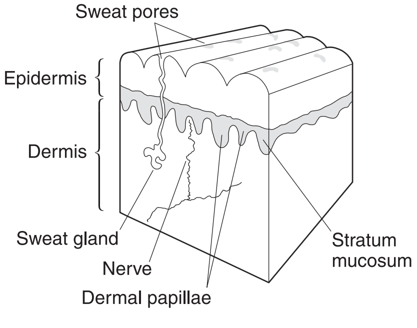


Figure 1 Structure of friction skin.

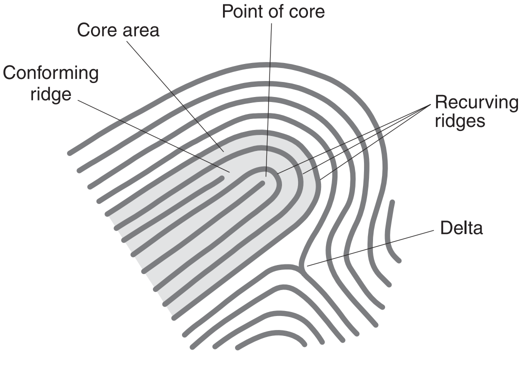


Figure 2 Elements of a fingerprint.

- **Delta:** When ridges converge on a point from three directions, the point is referred to as a triradius or delta. Arches and some tented arches lack deltas.
- **Core:** The central area of the pattern is referred to as the core and contains the point of core, all of the recurving ridges and the ridges which conform to them. The precise point of core is determined differently for each pattern type.

Fingerprint patterns

There are eight fingerprint pattern types.

- **Arch:** Ridges lying one above the other in an arching configuration. Arches lack both recurves and deltas (**Fig. 3A**).
- **Tented arch:** Arching formation with at least one upthrusting ridge at right angles to ridges above it (Fig. 3B).
- **Ulnar/radial loop:** Contains at least one free recurving ridge and one delta. In ulnar loops, recurving ridges originate from and return to the little finger side of the hand, named after the ulna forearm bone. In radial loops, recurving ridges originate from and return to the thumb side of the hand, named after the radius forearm bone (Fig. 3C).
- **Whorl:** Contains at least one free recurving ridge

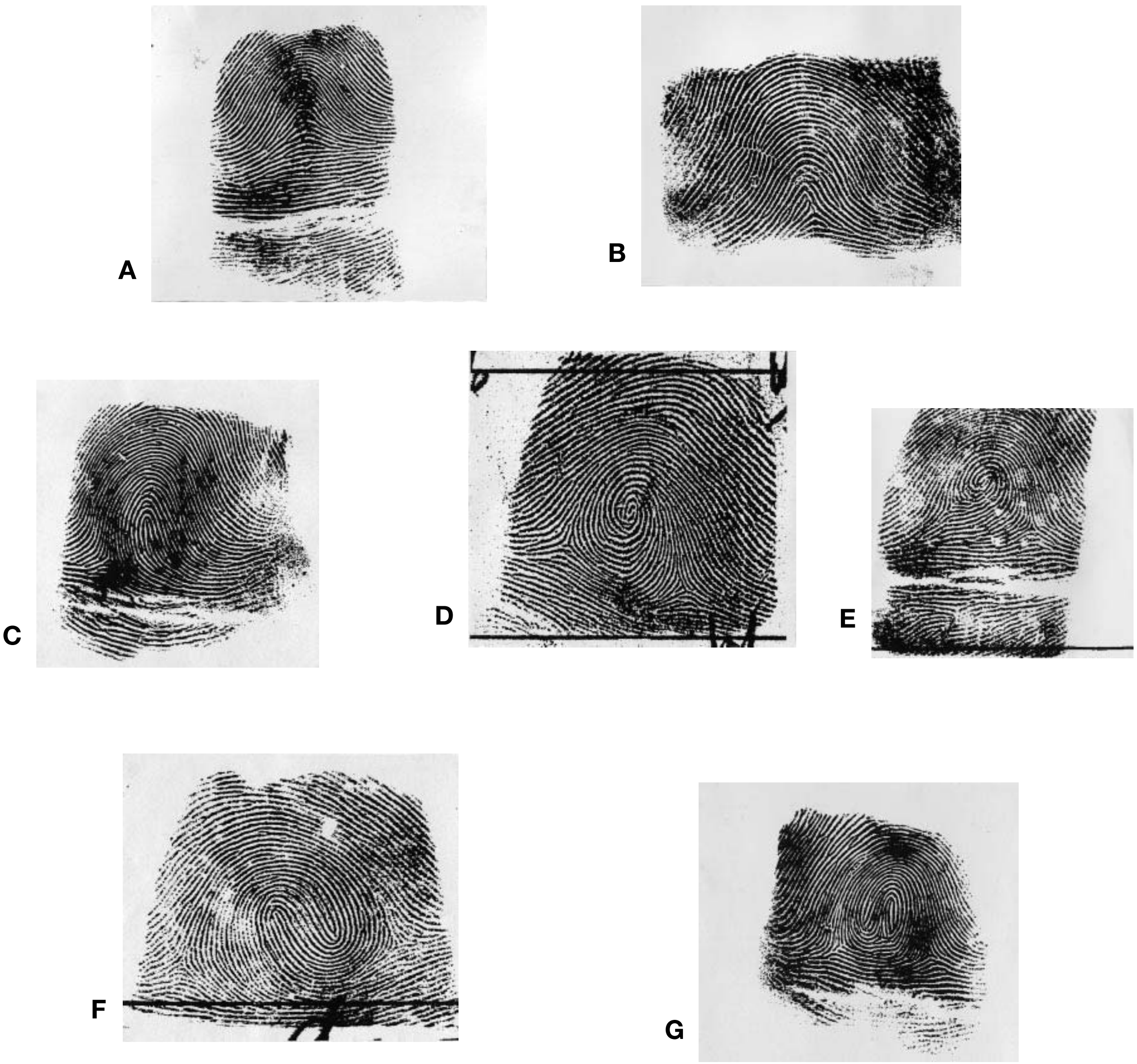


Figure 3 Fingerprint patterns. **A**, Arch; **B**, tented arch; **C**, ulnar/radial loop; **D**, whorl; **E**, central pocket loop; **F**, double loop; **G**, composite.

and two points of delta. A straight line between the points of delta will cut across at least one of the ridges revolving around the core (Fig. 3D).

- **Central pocket loop:** Contains at least one free recurving ridge and two points of delta. None of the ridges revolving around the core will be intersected by a straight line between the points of delta (Fig. 3E).
- **Double loop:** Contains two separate loop formations and two points of delta (Fig. 3F).
- **Composite:** Contains at least two different patterns, other than the basic arch (Fig. 3G).
- **Accidental:** Contains two deltas, one relating to an upthrust and the other relating to a recurve. Unlike the composite, the two pattern components in the accidental are amalgamated. In fingerprint pattern classification, no distinction is made between composite and accidental.

Loops, whorls, arches, indeed any variation from parallel ridges, may be found on any area of friction skin.

Ten print classification

In order to appreciate the value of classification, one might imagine a large reference library in which all the books are filed by the last name of the author. Anyone wishing to locate a book by subject alone would be faced with an impossible task. Library filing systems, however, enable one to find a book quickly, without knowing the author's name. Similarly, using the Henry Classification System, a 'John Doe' set of fingerprints could be searched and located in the file. Positive identification of criminals, with or without reliable names, had entered a new era.

In the Henry System, sets of fingerprints were given a primary classification, based on the distribution of patterns in the fingers. There are 1024 possible primary segments in the database, and further extensions were achieved by assessing the specific features of each digit. In order to conduct a search, subject fingerprints were classified in exactly the same manner as those in the database. Only the segment bearing the same classification required individual comparison to determine identity.

Single fingerprint classification

The Henry System allowed for the filing and searching of entire sets of fingerprints but was not useful in the search of individual 'cold' crime scene impressions. The Battley System was developed to address this need. At time of arrest, an extra set of fingerprints was taken on a serrated form. Each digit was classified, first by pattern type, and then by other extensions. The individual digits were then separated and filed. A crime scene impression was examined to determine the most probable digit(s), classified and searched. Several factors, however, have limited the speed and accuracy of the manual classification systems, as the databases grew in size.

1. Ulnar loops and whorls occur far more frequently than the other pattern types. Ten print classifications composed of these common patterns resulted in common classifications, and therefore comprised a large part of the collection.
2. As the size of a collection became more cumbersome, searches became more time consuming.
3. Many impressions do not fall exclusively into one of the pattern types. Depending on the amount of ink and pressure exerted during recording, an ulnar loop may resemble a tented arch or even a central pocket loop. It became necessary to 'reference' such impressions for a possible alternative pattern type, which often changed the primary classification.
4. Manual classification is also affected by individual interpretation. The technician making the search may not arrive at the same conclusions as the person(s) who classified the database.

The difference between fingerprint classification and identification must be clearly understood. Classification of fingerprints allows large segments of the database to be eliminated prior to the comparison process. It places a small number of possible matches in front of a technician for the final determination of identity or nonidentity. Regardless of the classification and search methods used, however, final comparison and identification is achieved by a fingerprint specialist.

For decades, manual fingerprint systems, including the Henry and Battley systems, provided a vital search tool in the identification of criminals from large fingerprint collections. They have been superseded, however, by automated techniques which excel in both speed and accuracy.

Identification

Philosophy

Fingerprint identification is based on two scientific axioms.

First, fingerprint detail is immutable. The relationship of ridge characteristics does not change from before birth until after death, except in size as a child reaches adulthood, and through injury, disease and decomposition. There is no arbitrary appearance or disappearance of the minutiae during life, and due to the thick, tough nature of friction skin, it is often the last of the soft tissue to decompose after death.

Second, there are no duplications within Nature. Every area of friction skin is unique to that individual on which it appears, to the exclusion of all others, living or dead. This position is, of course, unprovable in the conventional sense. No one has ever, or will ever have the capability to examine the entire friction skin of every human being who has ever lived. However, the overwhelming weight of precedent is on the side of this premise. Fingerprints have been in use for identification purposes for approximately 100 years. In that time no area of friction skin from one individual has ever been found to be identical to any other area, recorded either from the same individual or another.

The individuality of natural things is not limited to fingerprints. Some examples are more noticeable, including the stripe pattern on zebras and the striations on conch shells. Forensic studies have included the use of lip prints, ear prints, flexion creases and muzzle prints of cattle.

Identical twins

The most similar beings on earth are monozygotic (one egg) twins. They are the result of a single fertilized egg which subsequently divides to form two separate beings. The interval between fertilization and division can vary considerably. A zygote remaining as one entity for a longer period of time will undergo more common development before division, and the resulting twins will be even more alike. They are referred to as 'mirror image' twins. If the zygote divides after too much development has occurred, an incomplete separation (Siamese twins) will result. Fingerprints are formed randomly during the fourth fetal month, long after separation has occurred.

Current DNA technology is unable to differentiate between identical twins. Fingerprint patterns of identical twins may be strikingly similar, but the configuration and sequence of minutiae are as individual and unique as those of any two unrelated beings. It is significant that identical twins routinely do not share even the same patterns on their fingers. **Fig. 4** reveals the similarities and differences between identical triplets, the result of an egg dividing, followed by one of the twins dividing again.

Figure 4 Thumbprints of identical triplets.

Scars

Damage to the friction skin can be characterized as temporary or permanent. In the case of a minor injury such as a blister or paper cut, only the epidermis will be affected, and it will heal without scarring. If the injury extends to and includes the dermal layer where the fingerprint pattern resides, a permanent disruption in the fingerprint will result. At the conclusion of the healing process, this disruption or scar will be as permanent as the minutia configuration itself, and subject only to the same variations in appearance. As such, scars constitute a valuable addition to the identification process. The two fingerprints in **Fig. 5**, which were recorded 22 years apart, illustrate the persistency of permanent scars.

The comparison process

Fingerprint ridges routinely come to an end between two other ridges, or split into two (bifurcation). These disruptions are referred to as ridge characteristics, or minutiae, and occur on all friction skin (**Fig. 6**). There exists no common agreement as to the number and terminology of minutiae. Although many composites occur, there are only three basic types of minutiae; ridge endings, bifurcations and ridge dots. All other minutiae are combinations of these basic types. There are considerable differences in the rarity of minutiae, however, and commensurate weight must be given to this factor in the comparison process.

A trained expert concludes that two fingerprints have the same origin when he/she is satisfied that the sequence of minutiae is in continuous agreement and identical sequence. If the axiom regarding the individuality of ridge detail is accepted, it is not logical to accept that two impressions from different sources can have any true agreement of characteristics. The inability to distinguish between two impressions due to lack of sufficient and clear recording, however, is not only logical but routinely encountered. Also, no two impressions, whether they be latent, inked or any combination of these, will contain the same, or the same number of ridge characteristics. In the majority of cases, more characteristics will be present in an inked recording, although the reverse occasionally occurs. A comparison is therefore based on the characteristics common to both (**Fig. 7**).

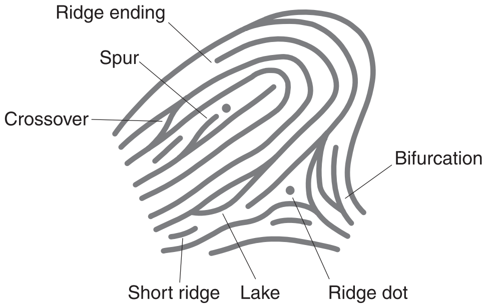


Figure 6 Ridge characteristics.

Poroscopy

The placement, frequency and configuration of the pores on the friction ridges is unique and potentially equal in identification value to the ridge endings (**Fig. 8**). Due to their small size, however, their appearance in a fingerprint depends on the sensitivity of

Figure 5 Scarred fingers recorded 22 years apart.

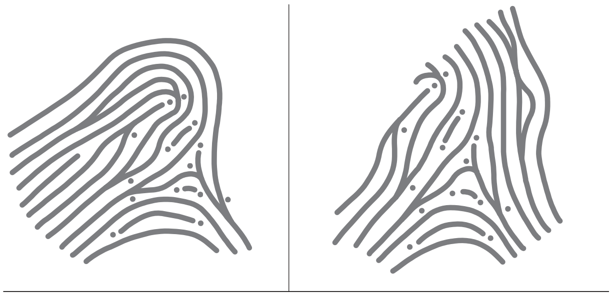


Figure 7 Agreement of ridge detail in sequence.

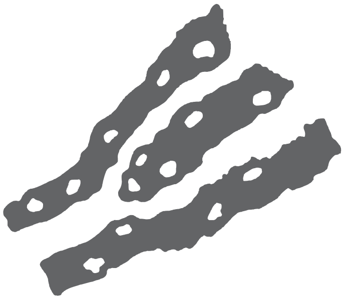


Figure 8 Sweat pores as they appear on ridges.

the recording technique, and they currently play a limited role in the determination of identity.

Size

Size relationships can play no part in determining fingerprint identity, except that ridge characteristics in two impressions from the same finger, will appear in the same general areas. The friction skin and the soft tissue behind it are flexible in nature. Inking, pressure and twisting, which cause distortion, are minimized when the print is professionally recorded. They vary greatly, however, in the everyday acts of lifting, pushing, grasping and sorting. Overlay techniques are therefore impractical in determining or displaying identity.

What do you see?

The comparison of fingerprints is obviously a visual exercise, but it involves much more than seeing. For example, a native Californian and a Bushman standing at a busy intersection in Los Angeles, will receive the same visual information but the manner and result of the information processing will vary greatly. Similarly, if the same two individuals are viewing an animal track in the Kalihari Desert, there will be a radical difference in information processing, although they both 'see' the same thing. In each case, only one observer possesses the skill and knowledge base to process the visual information successfully. This point must be clearly established as a prerequisite to understanding of fingerprint identification.

The attributes absolutely essential for the comparison and identification of fingerprints include:

1. A highly developed aptitude for recognition of spatial relationships;
2. Comprehensive knowledge of the anatomy, physiology and appearance of friction skin and recordings;
3. Extensive experience in examining and comparing a high volume of fingerprints.

Identification protocol

Determination of fingerprint identity requires adherence to a comparison protocol.

- The unknown impression is examined, and all minutiae are analyzed.
- The unknown impression is then compared to the known, to determine if a relationship exists.
- Weight is assigned in a comparison, not only to the number of minutiae in agreement, but also the rarity and clarity of those characteristics. Differences in appearance, due to recording technique, pressure and other factors, must be anticipated.
- If all characteristics common to both impressions are in agreement in identical sequence, with differences only in appearance, the only reasonable conclusion to be reached is that they were made by the same finger.

The issue of minimum number of points required for an identification has been continuously revisited by fingerprint specialists. It is generally agreed by all professional organizations that no set number of characteristics can be required for a conclusion of identity. This is a different issue entirely from jurisdictional policies which vary from country to country. Professional policy requires the independent comparison and endorsement in writing by a peer of the identification before positive findings are released.

Findings

Most fingerprint specialists believe strongly that only three opinions regarding the identity of the fingerprint should be volunteered.

1. Positive certainty: the fingerprint was made by the person indicated to the exclusion of every other person, living or dead.
2. Negative certainty: the fingerprint was not made by the person indicated.
3. Inconclusive: the impression lacks sufficient clear detail upon which to make a conclusion of either identity or elimination.

An opinion of possibility or probability is often solicited in court and must be addressed, but volunteering such opinions, or failing to qualify them by saying that the fingerprint could have been made by someone else, is generally viewed by the majority of professionals to be inappropriate.

There are a wide range of additional and valuable conclusions, however, which can be and should be drawn from the examination and comparison of fingerprints. These conclusions exploit the full expertise of the examiner and can extend far beyond the presence or absence, and the identity or nonidentity of a fingerprint. It is the responsibility of the technician that the maximum amount of probative information be extracted from the findings and relayed, either in the context of investigative leads or courtroom testimony.

These include, but are not limited to the following.

Location The exact location of a fingerprint on an exhibit or within a crime scene may determine its value to the investigation. For example, fingerprint evidence located at a robbery scene will be more significant if it is noted to be from an area not accessible to the public, or to refine it even further, not accessible to anyone but the owner.

Position The position and direction of a fingerprint identification may offer clues to the activity of the donor which may eliminate alternative explanations. An impression which ends abruptly along a fold line on a piece of paper can place that impression at a moment in time, and associate the person with a specific activity.

Similar fact The print of an accused trafficker repeated in the same position on many drug containers carries a more serious implication than simple possession of narcotics.

Absence of fingerprints

Another conclusion which is frequently reached in error by lay persons is that absence of a person's fingerprints on an object eliminates that person from having touched the object. If this were so, all of the following statements would be true.

1. Everyone always leaves identifiable fingerprints each time they touch something.
2. All surfaces are capable of receiving and retaining these prints, and are amenable to fingerprint examination.
3. Everyone's finger deposit always contains a sufficient quantity of the compounds targeted by the detection techniques in use.
4. The direct contact between finger and object was not physically obstructed by gloves, soil, tape or anything else.

In reality, fingerprint technicians constantly encounter obscured or fragmentary detail which indicates handling but is of no value in identification.

Duration of fingerprints

'How long do they last?' is one of the questions most frequently asked of fingerprint technicians. Fingerprints are often composed of very stable compounds. Unless they are exposed to extremes of heat, humidity and/or abrasion, they may persist indefinitely. The oldest substantiated impression was developed 42 years after it was deposited. There is currently no technology for dating fingerprints.

Conclusion

Fingerprints are the most powerful expression and frequent reminder of the fact that no two beings on earth, plant or animal, are exactly the same. Although these differences exist in all aspects of our bodies, we use our fingers to leave unique signatures throughout our environment, most often without the awareness that we have done so. The value of fingerprints in establishing positive and negative identity has been known and exploited for over a century. It remains today one of the most reliable and frequently used identification sciences to place a person in physical contact with a crime.

See also: **Cheiloscopy**. **Deoxyribonucleic Acid**: Parentage Testing. **Ear Prints**. **Fingerprints (Dactyloscopy)**: Chemistry of Print Residue; Visualization; Standards of Proof; Automated Methods, including Criminal Record Administration. **Pattern Evidence**: Bare Footprint Marks.

Further Reading

Almog J and Springer E (1996) The Ne'urim Declaration. *Proceedings of the International Symposium on Fingerprint Detection and Identification*, p. 21. Jersualem: Hemed Press.

Anthony CP and Alwyn IB (1976) *Structure and Function of the Body*, 5th edn. St Louis: CV Mosby.

Ashbaugh DR (1991) Ridgeology. *Journal of Forensic Identification* **41**:26–64.

Berry J and Banks P (1992) *Ridge Detail in Nature*. London: The English Fingerprint Society.

Canadian Police College (1977) *Fingerprints*. Ottawa: Royal Canadian Mounted Police.

Cherrell FR (1954) *The Fingerprint System at Scotland Yard*. London: HMSO.

Cummins H and Midlo C (1976) *Fingerprints, Palms and Soles*. South Berlin, MA: Research Pub.

FBI (1960) *The Science of Fingerprints*. Federal Bureau of Investigation, United States Department of Justice: US Government Printing Office.
Jungbluth WO (1989) Knuckle print identification. *Journal of Forensic Identification* **39**:375.
Montagna W and Parakkal PF (1974) *The Structure and Function of Skin*, 3rd edn. New York: Academic Press.
Olsen RD Sr (1977) *Scott's Fingerprint Mechanics*. Springfield, IL: Charles C. Thomas.
Ontario Police College (1986) *Fingerprint Identification*. Aylmer, Ontario, Canada: Ontario Police College.
Putter PJ (1982) An investigation to ascertain whether muzzle prints of cattle can be individualized. *Fingerprint Whorld*. **January**, p. 55.
Tuthill H (1994) *Individualization: Principles and Procedures in Criminalistics*. Salem, OR: Lightning Powder Company.

Sequential Treatment and Enhancement

S A Hardwick, D F Hewlett, T Kent and **V Sears**, Police Scientific Development Branch, Home Office, St Albans, Herts, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0519

Introduction

Although there are many chemical and physical techniques referred to in the literature for revealing and enhancing latent fingerprints there are only a relatively small number which are used widely and are most productive for operational use. The chemical methods for latent fingerprint detection all rely on reactions or interactions that occur between the applied reagent and components of latent fingerprints. Studies have shown that the distribution of chemical constituents varies widely between individuals, and for the same individual from day to day and hour to hour. The use of techniques in sequence will often therefore reveal more fingerprints than any single technique. Such sequential processing can be labor intensive and is sometimes only used in more serious crime investigations. Exhibits from minor crimes will often be treated with only the most cost-effective and straightforward process.

The careful choice of processes and their sequence can dramatically increase the number of fingerprints detected leading to greater numbers of identifications.

Fingerprint development techniques can be classified both by what they react with in the fingerprint residue, and by the types of surface on which they can be used. Since very little is known about the distribution of natural, or contaminating, components deposited in a fingerprint, the nature of the surface touched is usually a major determining factor in the selection of treatments. In this article fingerprint development techniques will be described for nonporous, porous and adhesive surfaces. Methods that can be used for enhancing fingerprints contaminated with blood and grease etc. are also discussed. Sequences of techniques for other surfaces such as metal, wood and plastic wrapping materials have also been determined.

Great care must be exercised in the packaging and handling of articles throughout the procedure, from retrieval at the scene of crime to final photography, to ensure that latent or developed fingerprints are not damaged. Those on nonporous surfaces are particularly easily damaged.

It is essential that all visible ridge detail is recorded photographically at each stage in a sequence of treatments since some fingerprints may be lost during subsequent treatments.

Factors Affecting the Choice of Development Technique

As well as considering the types of surfaces that are to be treated, it is important that the following points are considered when selecting a sequence of treatments.

1. The importance of the case and resources available.
2. Techniques should be used if possible in an order in which the effectiveness of methods used later in the sequence are not adversely affected by those used at the beginning of the sequence.
3. Any contamination on the surface that could help or hinder fingerprint development, e.g. blood, grease etc.
4. Sampling of biological material prior to treatment for fingerprints. If this is not possible due to insufficient deposit of material, specialized advice on the effects of fingerprint development techniques on DNA profiling should be obtained.
5. Samples of other trace evidence such as hairs or fiber should usually be taken prior to treatment for fingerprints but in a way that does not compromise

the fingerprint evidence. ESDA for developing impressions in paper must also be performed before chemical treatments for fingerprint development are applied.

6. The effects of exposure to the elements, especially exposure of the surface to rain or moisture may affect the choice of treatment sequence.
7. Full use is made of optimized reagent formulations and specialized equipment at every stage of examination for fingerprints.

Detailed information on the selection of latent fingerprint development techniques can be found in the *Manual of Fingerprint Development Techniques* listed in the bibliography. Flow charts in this publication have been developed to graphically represent the selection of processes for specific surfaces, an example of which is given in **Fig. 1**.

What follows here is a synopsis of the basic concepts that underpin the selection of techniques. Summaries of the compatibility of fingerprint devel-

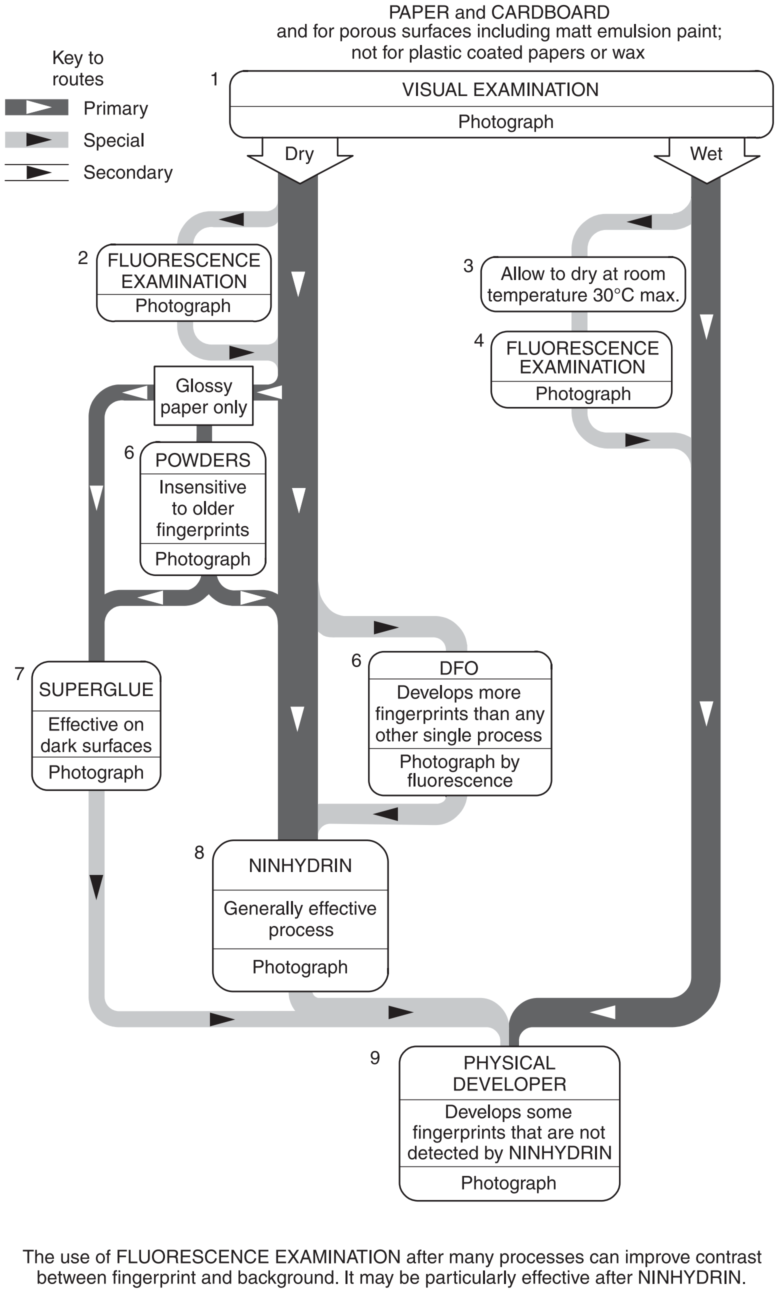


Figure 1 Process selection chart for smooth nonporous surfaces (From Kent T (ed.) (1998) *Manual of Fingerprint Development Techniques*, 2nd edn, London: Home Office.)

opment techniques with surface type, and with contamination from water, blood and grease etc. are given in **Table 1**.

Fingerprint Development: Nonporous Surfaces

Background

Many articles at scenes of crime fall into this category including glass, gloss-painted surfaces, metal and plastic. Fingerprints on nonporous surfaces can be very easily damaged as most of the trace residues deposited from the fingerprint ridges usually remain on the surface.

Development techniques

Visual examination Visual examination of any article may sometimes reveal latent fingerprints, particularly if the surface is smooth and clean. A variety of light sources should be used, e.g. oblique illumination. Any fingerprints should be photographed before proceeding with other techniques.

Fluorescence examination Examination of surfaces with a high intensity light source or laser, usually working in the near UV, blue or green parts of the spectrum and using barrier filters in front of the eyes, will sometimes reveal fingerprints either by the rarely observed fluorescence of naturally occurring components of the latent fingerprint, or more often by the

Table 1 Summary of the use of fingerprint development techniques

	Surface type			*Contamination*		
Process name	*Nonporous*	*Porous*	*Adhesive*	*Blood*	*Grease*	*Effectiveness on wetted surfaces*
Amido Black[a]	√	√		√ Preferred method		
DFO		√		√		× Ineffective
Fluorescence examination	√	√	√	√	√	√
Gentian violet[a]	√		√ Preferred method		√	√ Can be used after drying
Ninhydrin		√		√		× Ineffective
Physical developer		√	√	√	√	√ Sole method for wet paper
Powders	√			√	√	√ Can be used after drying
Small particle reagent	√					√
Sticky-side powder			√			√
Sudan Black[a]	√				√ Preferred method	√
Superglue	√		√			× Ineffective
Vacuum metal deposition	√					√ Can be used after drying
Visual examination	√		√	√	√	

[a] Can be used effectively at scenes of crime.

fluorescence of some contamination which may have been on the finger such as oil or grease. Fingerprints can also sometimes be visualized by a combination of light absorption by the fingerprint ridges and fluorescence of the background.

Vacuum metal deposition (VMD) Of the methods that are used to develop latent fingerprints on nonporous surfaces, this is the most sensitive, being capable of detecting monolayers of fat by sequential deposition of a very thin coating of thermally evaporated gold followed by zinc. Plastic wrapping materials up to about 1000 mm × 1000 mm and other small solid objects may be treated in specially modified industrial vacuum coating systems. VMD has the great advantage that it can develop fingerprints on surfaces that have previously been wet or even submerged in water for extended periods of time. If fingerprints are not revealed by VMD, superglue, powders or other techniques may be used subsequently.

Fingerprint powders Powdering, one of the oldest techniques for detecting fingerprints on smooth nonporous articles, is quite effective and widely used; many latent fingerprints are still found by this simple method. Although there are many powders described in the literature, those with a flat, flake like structure such as milled aluminum or brass, or molybdenum disulfide are significantly more sensitive and generally more effective at developing fingerprints on smooth, clean surfaces than the more traditional black or white powders which are of a more granular nature. The metallic flake powders are most effectively applied with a fine glass fiber brush. Other powders are often applied with soft animal hair brushes.

Surfaces which are dirty or contaminated will sometimes take up too much of the flake powders all over the surface so that granular black or white powders may be more suitable. Rough or grained surfaces may sometimes be treated using iron, cobalt or nickel-based powders used with a magnetic applicator although superglue fuming is probably to be preferred.

There are many fluorescent powders which may also be used in conjunction with a suitable light source. Powders may be used on surfaces which have been wet, but only after the articles have been thoroughly dried at temperatures not exceeding 30°C.

Superglue fuming Superglue fuming can be used on many nonporous surfaces and is particularly useful on surfaces such as rough or grained plastic surfaces which cannot be easily treated using VMD. This technique is particularly effective when performed under optimum controlled conditions (normally 80% relative humidity, ambient temperature and pressure). When superglue fuming is performed under these high humidity conditions fibrous polymer growth occurs across the ridge deposit and is believed to be initiated by water droplets present on chloride deposits in the fingerprint deposit. This polymer growth increases mechanical strength and robustness of the developed fingerprint and produces a very large surface area, this appears white but can be very effectively enhanced by means of a solution of fluorescent dye which is absorbed and may be visualized using fluorescence examination. Use of fluorescent staining techniques instead of direct photography of the white superglue deposit can double the number of fingerprints detected. Many dyes have been used but Basic Yellow 40 in ethanol is very effective and is believed to be of lower toxicity than some others.

Superglue fuming may be carried out at reduced pressure but fingerprints developed under these conditions have a smooth surface with less surface area, are more difficult to visualize and absorb less dye.

Superglue fuming can be used after VMD and powders but is ineffective if used after a water-based development technique, or on articles that have been wet.

Small particle reagent (SPR) SPR consists of a suspension of molybdenum disulfide suspended in an aqueous detergent solution. This is used by agitating the reagent mixture so that the particles are temporarily suspended in the liquid and holding the article to be treated under the liquid allowing the molybdenum disulfide particles to deposit onto the surface. The article is gently removed and drawn through a container of clean water. The molybdenum disulfide particles adhere to fats deposited in the fingerprint, producing a gray–black image. SPR is used in dishes or tanks for portable articles. SPR can also be used by spraying directly onto larger surfaces, however used in this fashion it is considerably less sensitive than the submersion method.

Treatments for nonporous surfaces which are contaminated with blood or grease are discussed in a later section.

Preferred sequence of techniques The order in which techniques are used will be determined by the following factors:

1. the importance of the case;
2. the size of the article;
3. whether the surface is rough or smooth;
4. whether the surface has been wet;
5. surface contamination.

When these factors are taken into account, sequences of treatments can easily be determined. For example, a plastic bag from a serious crime which is not contaminated with blood etc. might be treated by visual examination and fluorescence examination followed by VMD and possibly superglue fuming or SPR.

Fingerprint Development: Porous Surfaces

Background

The most common porous surfaces examined for fingerprints include paper, wall paper, cardboard and matt emulsion painted surfaces. The reagents used for these surfaces react either with amino acids, traces of fats and lipids or chlorides absorbed into the surface.

Development techniques

Visual examination Visual examination is less likely to reveal fingerprints on porous surfaces unless they have been contaminated with dirt, blood, grease or similar materials.

Fluorescence examination Examination of such surfaces will sometimes detect fingerprints either by the rarely observed fluorescence of naturally occurring components of the latent fingerprint or fluorescence of some contamination which may have been on the finger such as oil or grease. Fingerprints can also be visualized on fluorescent surfaces when the fingerprint deposit absorbs light; again such fingerprints should be photographed before proceeding.

1,8-Diazafluoren-9-one (DFO) The most sensitive reagent currently available for detecting fingerprints on porous surfaces is DFO. This compound reacts with amino acids deposited in the fingerprints to produce a faintly colored but intensely fluorescent compound which can be easily photographed. Since amino acids are soluble in water, neither DFO nor ninhydrin can be used to treat porous surfaces which have been wet.

Articles are dipped into a solution containing DFO, acetic acid and an inert carrier solvent and then heated in a dry oven at 100°C for 20 min. If humidified conditions are used for DFO-treated articles, no fingerprints are detected. DFO is not routinely used at scenes of crime because it is difficult to generate sufficient heat in a room to produce an acceptable rate of fingerprint development although local heating can be used for small areas.

Ninhydrin This is a widely used process which also reacts with amino acids and produces a purple colored product. Articles are dipped into a solution of ninhydrin and acetic acid in an inert carrier solvent and then heated in a humid atmosphere for rapid fingerprint development, the optimum conditions are 65% relative humidity and 80°C for 3–4 minutes. (If these same humidification conditions are used for DFO-treated articles, no fingerprints are detected.) The developed purple-colored fingerprints can be easily recorded using conventional photography, a green filter can improve contrast.

Ninhydrin can be very effectively used at scenes of crime with the same formulation being brushed onto the surfaces. Sealing the room and raising the temperature with domestic heaters will assist fingerprint development but development will take several days. The time taken for this technique varies widely because humidity and temperature cannot be optimized.

In large-scale comparisons it has been shown that DFO develops about 60% more fingerprints than are found using ninhydrin. However, when ninhydrin is used after DFO, a further 10% of fingerprints are developed compared to instances where DFO is used alone. DFO is ineffective in developing latent fingerprints if it is used after ninhydrin.

When ninhydrin-developed fingerprints are developed on dark nonfluorescent backgrounds or when the background is a very similar color to the fingerprint zinc toning can be used to enhance visualization. The fingerprint is treated with zinc chloride solution and gently heated to produce a fluorescent fingerprint which can then be photographed.

Powders Smooth papers may be treated with some types of black or magnetic powder although these will usually only detect heavier or more recent fingerprints.

Superglue fuming This may be used on some smooth surfaces such as cigarette packets but enhancement with fluorescent dyes may be difficult due to absorption of dye by the background.

Physical developer (PD) In cases when a porous surface has been wet, the only available technique for detecting fingerprints is PD. This reagent is an aqueous solution of silver nitrate containing a Fe(II)/Fe(III) redox couple and two detergents. The exact reaction mechanism is unknown although it may detect intercalated sodium chloride molecules present in the small amounts of fat deposited in the absorbed fingerprints. The developed fingerprints are gray–black in color and can be recorded using conventional photography. PD is impractical to use at scenes of

crime. Fingerprints developed by the process may be enhanced by radioactive labeling or mapping of the silver deposit using X-ray radiographic techniques.

Treatments for porous surfaces which have been contaminated with blood or grease are discussed in a later section.

Preferred sequence of techniques The preferred sequence of fingerprint treatments on porous surfaces would be: visual examination; fluorescence examination; powders (where applicable); superglue fuming (where applicable); DFO; ninhydrin; physical developer. The actual order to be used in a particular case will be determined by:

1. the importance of the case;
2. whether or not the article has been wet;
3. whether the paper surface is very smooth.

If the article has been wet, DFO and ninhydrin are omitted and physical developer used by itself.

If the paper has a glossy finish, the use of powders or superglue fuming described in the section on nonporous surfaces may be more appropriate. Good examples of this sort of surface are cigarette packets and glossy magazines.

Fingerprint Development: Adhesive Surfaces

Background

Surfaces in this category include the adhesive sides of sticky tapes and labels. There are three general techniques that are used to develop latent fingerprints on these surfaces: gentian violet, superglue fuming and powder suspensions. No significant comparative trials are described in the literature. The gentian violet technique has been used for many years but other techniques are becoming widely used. Physical developer may also be used on certain types of paper tapes.

Development techniques

Gentian violet (crystal violet) The most effective gentian violet formulation (also known as crystal violet) is a solution of crystal violet (CI 42555), phenol and ethanol in water. Adhesive tapes are drawn across the surface of the solution two or three times so that the dye is absorbed by the fingerprint ridges. Other adhesive surfaces to be treated with gentian violet can be floated on top of, or held in contact with the solution for ease of processing. Once treated, the surfaces are washed in slowly running water. Fingerprints on lightly colored surfaces will be visible to the naked eye and can be recorded using conventional photography. Fingerprints on black or dark-colored surfaces will not be immediately visible. This problem can be easily overcome by transferring the developed fingerprints onto photographic paper. Once this has been done it is important to remember that the fingerprints will have been laterally reversed. It is important to bear in mind that fingerprints on the adhesive side of tapes may already be laterally reversed before transfer onto photographic paper.

Superglue fuming Superglue fuming of adhesive surfaces is accomplished in the same way as on nonporous surfaces. Again, fluorescent dyeing and subsequent fluorescent examination may provide significant enhancement although background uptake can be a problem.

Powder suspensions Powder suspensions such as 'Sticky-side Powder®' can also be used to develop fingerprints on adhesive tapes. A concentrated suspension of the powder is applied to the adhesive side of the tape, either by dipping or careful brushing. The solution is left in contact with the tape for 10–15 s and then washed off using slowly running water. Fingerprints are seen as gray–black images which can be recorded using conventional photography. Precise treatment times and reagent concentrations have not been determined or reported to date.

Preferred sequence of techniques Little is known about the relative effectiveness of the three techniques above but the use of powder suspensions or GV will render superglue ineffective as an after-treatment as both of the reagents are water based.

Fingerprint Development: Blood Contamination

Background

It is generally advisable that samples of blood are taken for DNA and other analyses before fingerprint treatments. If this is not possible, specialized advice should be sought to determine which fingerprint development techniques are compatible with DNA profiling; the latter is a rapidly developing area of technology and compatibility with fingerprint development techniques is likely to need reviewing.

None of the techniques that are commonly used to enhance bloody fingerprints are specific for blood and therefore cannot be used as presumptive tests for blood.

There are many techniques used for blood splatter imaging that rely on the iron centers in hemoglobin to

catalyze chemical reactions which result in color changes or chemiluminescence. Some of these techniques can be adapted for use in enhancing fingerprints in blood although few have been fully evaluated for this purpose.

Development techniques

Fluorescence examination An effective method for detecting bloody fingerprints is by using fluorescence examination to excite fluorescence of the substrate. Blood itself is not fluorescent but absorbs light. Consequently, if the background surface can be made to fluoresce, the fingerprint can be viewed as black ridges on a white background. Fluorescence examination should be the first technique used to enhance latent fingerprints as it is simple to use and nondestructive provided significant heating is avoided.

Amido black The most commonly used dye for enhancing bloody fingerprints is Acid Black 1 (CI 20470) which is generally referred to as amido black. This is a general protein stain which binds to the proteins in blood plasma and to blood corpuscles; amido black is not a specific blood stain.

Amido black is commonly used as a solution in methanol. However, if a surface has to be treated that would be adversely affected by methanol such as some types of plastic, a water-based formulation may be substituted. Fingerprints developed by amido black are dark blue/black in color and can be recorded using conventional photography.

DFO and ninhydrin DFO and ninhydrin can also be used to develop bloody fingerprints on porous surfaces. These reagents react with amino acids in the blood residue to give fluorescent or colored fingerprints, respectively. As stated in the section on porous surfaces, DFO must be used before ninhydrin for fingerprint development to be effective.

Physical developer (PD) Very little information is available on the relative effectiveness of this technique for enhancing bloody fingerprints compared to amido black, DFO or ninhydrin.

Powders Powders can also be used to visualize bloody fingerprints when only part of the ridge detail is in blood; carefully applying powder will develop the ridges which are not contaminated with blood.

Preferred sequence of techniques The preferred sequence for enhancing bloody fingerprints on porous surfaces is as follows: fluorescence examination, DFO, ninhydrin, amido black then physical developer.

With one exception, the sequence for bloody fingerprints on nonporous surfaces is simpler as neither of the amino acid reagents can be used: fluorescence examination, amido black and PD. When a fingerprint is only partially in blood, fingerprint powders may also be used.

The exception is where a bloody fingerprint or partially bloody fingerprint is on a black, nonfluorescent surface. Under these circumstances both fluorescence examination and amido black are ineffective, so that the only effective process is ninhydrin followed by zinc toning. In this case the ninhydrin reacts with amino acid and/or protein residues in the blood to produce Ruhemann's purple, which may not be visible. The Ruhemann's purple is then converted into a fluorescent product by further treatment with zinc chloride rendering the fingerprint recordable using fluorescence examination techniques. Other fluorogenic reagents that detect amino acid residues in fingerprints such as DFO have not proven effective in enhancing fingerprints heavily contaminated with blood on nonporous, nonfluorescent surfaces.

Fingerprint Development: Grease Contamination

Introduction

Fingerprints in grease and other types of surface contamination can be enhanced using a number of techniques such as sudan black, gentian violet and occasionally physical developer. Other types of contamination can also be examined using these techniques including sticky deposits on the outside of drink cans and milk cartons etc.

Sudan black Sudan black uses an ethanolic solution of a dye called Solvent Black 3 (CI 26150) to stain the greasy fingerprint ridges. The surface under examination is treated with the solution for 2 min either by immersing the article in the solution or by floating the article on top of the solution. The surface is then gently rinsed under slowly running tap water and allowed to dry. Fingerprints are recorded using conventional photography.

Physical developer and gentian violet When enhancing greasy fingerprints on porous surfaces physical developer can also be used. On nonporous surfaces, gentian violet can be used to effectively

develop fingerprints. Little information is available on the relative effectiveness of either of these techniques.

See also: **Crime-scene Investigation and Examination**: Fingerprints. **Fingerprints (Dactyloscopy)**: Chemistry of Print Residue; Visualization. **Photography and Digital Imaging**: Overview.

Further Reading

Almog J (1987) Reagents for chemical development of latent fingerprints: vicinal triketones – Their reaction with amino acids and with latent fingerprints on paper. *Journal of Forensic Science* **32**:1565–1573.

Almog J and Springer E (eds) (1995) *Proceedings of the International Symposium on Fingerprint Detection and Identification*. Jerusalem: Israel National Police.

Burns DS (1994) Sticky-side powder: the Japanese solution. *Journal of Forensic Identification* **44**:133–138.

Gower JF (1983) *Friction Ridge Skin*. New York: Elsevier Science.

Hardwick SA, Kent T and Sears VG (1990) *Fingerprint Detection by Fluorescence Examination. A Guide to Operational Implementation*. London: Home Office.

Hauze DB, Petrovskaia O, Taylor B, Joullié MM, Ramotowski R and Cantu AA (1998) 1,2-Indanediones: new reagents for visualising the amino acid components of latent prints. *Journal of Forensic Science* **43**:744–747.

Kent T (ed) (1993) *Scene of Crime Handbook of Fingerprint Development Techniques*. London: Home Office.

Kent T (ed) (1998) *Manual of Fingerprint Development Techniques*, 2nd edn. Sandridge: Home Office.

Knowles AM (1978) Aspects of physiochemical methods for the detection of latent fingerprints; review. *Journal of Physics E: Scientific Instruments* **11**:713–721.

Lee HC and Gaensslen RE (eds) (1991) *Advances in Fingerprint Technology*. New York: Elsevier Science.

Maehly A and Williams RL (eds) (1988) *Forensic Science Progress*. Berlin, New York: Springer-Verlag.

Menzel ER (1980) *Fingerprint Detection with Lasers*. New York: Marcel Dekker.

Olsen Sr RD (1978) *Scott's Fingerprint Mechanics*. Springfield, IL: CC Thomas.

Thomas GL (1978) The physics of fingerprints and their detection. *Journal of Physics E: Scientific Instruments* **11**:722–731.

Standards of Proof

C Champod, Institut de Police Scientifique et de Criminologie, University of Lausanne, Lausanne-Dorigny, Switzerland

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0521

Introduction

Papillary surfaces are covered in fine lines, or 'ridges' arranged generally in patterns such as loops, whorls and arches. The ridges form characteristics called minutiae such as bifurcations, ridge endings (or combined minutiae) and also microscopic features (pores, ridge edges and structures). These detailed features are the consequence of a morphogenesis so sensitive to outside influence that they are unpredictable, unlike general patterns, in their position and their form. The variability is such that even monozygotic twins have different fingerprints in this respect.

Following a comparison between a mark (recovered for example in association with a crime) and a control print (associated with a suspect), suppose that the examiner observes a complete agreement between the mark and the print without any significant discrepancy. The question is then often expressed as follows: 'how many similarities are required to confirm the identification?' or 'what is the evidential value of a partial print with such similarities?'

The aim of this article is to try to answer these questions, to address the question of the 'standard of proof', by reviewing international views and practices.

Historical Milestone

The first rules establishing the minimum number of minutiae necessary for fingerprint identification can be attributed to Edmond Locard, in 1911–1912. He suggested a tripartite rule, which followed from the discovery of poroscopy. This can be summarized as follows.

1. If more than 12 concurring points are present and the fingerprint is sharp, then the certainty of identity is beyond debate. (The imperative requirement for the absence of significant differences is implicit.)
2. If there are 8–12 concurring points, then the case is borderline and the certainty of identity will depend on:
   (a) the sharpness of the fingerprint;
   (b) the rarity of its type;
   (c) the presence of the center of the figure (core)

and the triangle (delta) in the exploitable part of the print;
(d) the presence of pores;
(e) the perfect and obvious identity regarding the width of the papillary ridges and valleys, the direction of the lines, and the angular value of the bifurcations.
In these instances, a certainty can only be established after discussion of the case by at least two competent and experienced specialists.
3. If a limited number of characteristic points are present, the fingerprint cannot provide certainty for an identification, but only a presumption proportional to the number of points available and their clarity.

Locard based his tripartite rule on various sources of information: the discovery of poroscopy, the practice gathered by the identification bureaux around the world and statistical evaluation.

This approach persists throughout the extensive writings of Locard. His considerations (principally the first two) were largely taken up by the most eminent dactyloscopists or criminalists of the first half of this century.

Current Views and Practice

On the fringe of the theoretical view expressed above, the practice has led to various positions. The different positions that currently exist were not established without controversy after the views expressed by Locard.

Predetermined minimum number of minutiae

The majority of European dactyloscopists have favored a purely quantitative approach, leaving qualitative aspects in the background, by fixing a numerical standard – a minimum number of minutiae necessary to establish identification. The numerical standard represents a lower limit; above this value, the identification is beyond doubt regardless of the type of minutiae that are found. The interpretation of the concept of numerical standard may vary from agency to agency; some positions are summarized in **Table 1**.

Despite the systematic use of a numerical standard, various countries (e.g. Finland, Greece, Holland, Israel and Portugal) have developed methods to bypass the rigid number when certain characters (visibility of pores, ridge structures or rare combinations of minutiae) are observed in the comparison. The adoption of a range from 8 (or 10) to 12 points is a way to relax a rigid threshold. In 1983, an addendum was made to the '16 point standard' in the UK, which stated that in extremely rare cases, an expert with long experience and high standing in the profession can make an identification that does not meet the nationally accepted standard.

No predetermined numerical standard

Before the 1970s, American fingerprint examiners followed Locard's view. As a result, the 12-point rule was generally respected and below this threshold qualitative factors in the comparison were taken into consideration. In 1970 a commission of experts from

Table 1 Principal positions regarding numerical standards adopted in some countries

Country	*Numerical standard*	*Origin (when known) and specificity*
Italy	16–17	Probabilistic calculation by Balthazard dating back to 1911. The minimum standard is expressly mentioned from a jurisprudence referring to Balthazard's work. The jurisprudence dated back to 1954, but has been confirmed in 1959 and 1989.
Germany, Sweden, Switzerland	8–12	In agreement with Locard even though in practice there is a clear tendency to respect a '12-point' rule.
UK (before 2000)	16	The origins of this standard (adopted by New Scotland Yard in 1924) date back to an erroneous interpretation of a document published by Bertillon in 1912. The numerical standard is virtually impossible to circumvent. Its purpose is to guarantee a high level of quality and faultlessness in the matter of fingerprint identifications; it is a means of assuring the weakest link in the chain. In this view, dactyloscopic evidence must remain an 'absolute' form of proof.
Belgium, Finland, France, Holland, Israel, Greece, Poland, Portugal, Romania, Slovenia, Spain, Turkey, South American countries	12	Number probably derived from the first rule of Locard.
Russia	7	According to an Interpol survey among European countries.

the International Association for Identification (IAI) was established to study the question of the pertinence of a fixed numerical standard for dactyloscopy. The work of this committee (for 'standardization') resulted in a number of articles and reports tracing the state of empirical and scientific knowledge.

Apart from the fact that the justification of any number could not be based on statistical researches pertaining to fingerprints, the most powerful argument against any numerical standard derives from knowledge of the morphogenesis of the papillary lines. The development of papillary lines on corresponding surfaces is initiated (for the fingers) after 7 weeks of fetal life with the growth of volar pads on a papillary area. After 10 weeks, as the volar pads regress, primary ridges begin to grow in the dermis followed by secondary ridges. The various stresses involved in this process (regression of the volar pads, development of size, meeting of multiple development fronts) induce a stochastic formation of minutiae (in terms of form and positioning). Around week 25, the development of papillary lines is completed on the dermis and is projected onto the epidermis of the skin. The fact that the complete pattern is secured on the dermis explains the durability of the fingerprints. From that moment, a final differentiations occurs on the papillae which dictates the form of the pores and ridge edges. This final stage also produces a strictly individual pattern.

Hence, it was clearly not legitimate to reduce the fingerprint individuality to the minutiae as is suggested with any numerical standards. The number of specific features is much broader than minutiae alone. The nature of the papillary individuality prevents the adoption of any predefined number of ridge characteristics necessary (without significant differences) for identification.

Following such arguments, the following resolution was adopted by the International Association for Identification in 1973: 'The International Association for Identification, based upon a 3-year study by its Standardisation Committee, hereby state that no valid basis exists for requiring a predetermined minimum number of friction ridge characteristics that must be present in two impressions in order to establish positive identification.

This resolution is also endorsed by practical observations.

- Some minutiae are rarer than others (for example, a double bifurcation in the periphery of the pattern is six times less frequent than two separate bifurcations). Hence, counting minutiae up to a predefined number without taking into account their relative and distinct values is not adequate.
- The absence of any minutiae; from a papillary surface is as relevant as their presence. It would be very exceptional to observe a core area or a delta area where the ridges do not display any minutiae.
- The pore structure and ridge edge structures are individual and, when visible, contribute to the identification decision. Numerous cases of identifications based on poroscopy and edgeoscopy (i.e. ridgeology) are documented in the scientific literature.

It has been accepted that the concept of identification can not be reduced to counting fingerprint minutiae; each identification represents a unique set of circumstances and the identification value of concurring points between two fingerprints depends on a variety of conditions that automatically excludes any minimum standard. When a fingerprint expert concludes an identification, he reaches a decision threshold. This threshold may be a number of concurring points or, in addition to this quantitative element, consideration of qualitative factors such as the rarity of the general pattern, the type of points observed, and the relative frequencies of these points. In this way, the identification process is a global assessment which balances both quantitative and qualitative aspects. Following this resolution, different minima were instituted at the level of the training of American dactyloscopists. A technical working group driven by the FBI on friction ridge analysis (SWGFAST) has reviewed these guidelines.

In 1995, a conference meeting on fingerprint detection techniques and identification at Ne'Urim (Israel), unanimously approved a resolution, a slight variation of the IAI 1973 resolution. The Ne'Urim declaration states that: 'No scientific basis exists for requiring that a predetermined minimum number of friction ridge features must be present in two impressions in order to establish a positive identification.' Various countries are or will follow the IAI as indicated in **Table 2**.

It must be noted that the work done by the English committee has been a determinant in the various moves towards the adoption of the IAI resolution. In 1989, a collaborative study among 130 fingerprint examiners in England and Wales was conducted. Each participant was sent 10 mark/print comparisons (9 associations and one exclusion) and was asked to mark the number of concordant minutiae and to express an opinion with respect to the identification. No misidentification was reported. But, when examining genuine pairs of prints, examiners varied widely in the number of points of comparison they found. In the case of one print it varied from 11 to 40. In other words, some examiners would not have gone

Table 2 Main positions regarding the abandonment of a numerical standard

Country	*Numerical standard*	*Origin (when known) and specificity*
USA and Canada	No	Since 1973, following the IAI resolution.
Norway	No	The other Scandinavian countries (Sweden, Finland) are now discussing a potential move towards the abandon of any numerical standard.
UK (before 2000)	No	In 1988, a committee was named by ACPO (Association of Chief Police Officers) and the Home Office to undertake a review of the origin and relevancy of the 16 point standard. Following this report, the abandonment of the numerical standard was recommended.
Australia	No	In 1996, the '12 point rule' adopted in 1975 was abandoned in favor of the IAI resolution. A working party was formed to examine various issues such as a National Accreditation Board (education and competency assessment) and other quality assurance issues.
Switzerland (long-term objective adopted by heads of fingerprint bureaux in 1997)	No	Taking advantages of the debate in the UK and Australia and following the Ne'Urim declaration, a committee has undertaken to manage the change towards a nonnumerical practice.

to court with this pair of impressions although most would have done so. An identical survey in Switzerland in 1997 led to analogous results. These studies led to the following conclusions.

- The dogma adopted by some fingerprint examiners that fingerprint identification is an 'exact' science is a misconception. The essence of science is inductive inference. Inference is a mental process that cannot be exact (or deductive). Fingerprint identification is scientific in that sense.
- The precision implied by any number (12, 16, etc.) is also a lure. The determination of individual minutiae or features is highly subjective.
- As a means of achieving quality, a numerical standard is poor. The way forward is to concentrate on professional standards rather than on rules about numbers of minutiae. A scheme of quality management should be instituted including training, certification testing, performance testing, file audits and blind trials.

Range of possible conclusions in the fingerprint field

Most contemporary fingerprint experts refuse to give qualified opinions on fingerprint comparisons that do not meet the minimum requirements for an identification. From this standpoint, it is not authorized to speak of 'possible' or 'probable' identifications, for example.

The idea that dactyloscopy could provide a presumptive link preceded Locard, notably in the casework dating back to the beginning of the twentieth century by Reiss at the Institut de Police Scientifique et de Criminologie (IPSC) in Lausanne, Switzerland. At present, very few identification bureaux in Switzerland leave this possibility open. However according to a recent European survey, few other countries are using fingerprint evidence as corroborative evidence. For example in Belgium where there are between 8 and 12 points of agreement between a questioned fingermark and a fingerprint (without discrepancies), it is stated that 'the fingerprint could be the same as the fingermark being examined'.

Therefore, apart from a few exceptions, Locard's third directive was totally forgotten as the use of dactyloscopy became widespread, thus giving dactyloscopy its very distinctive character compared to other forms of transfer evidence. In practice nowadays, a fingerprint is proof of identification or exclusion or it has no evidential value. In most countries, the probabilistic aspect has been excluded from dactyloscopy, removing the possibility for a fingermark to constitute corroborative evidence for an identification.

Steinwender, a German fingerprint expert, was the first to publish his refusal to consider dactyloscopic evidence as anything but absolute proof. This point of view has largely been followed by the whole profession. This is sometimes described as the 'positivity' of the fingerprint field.

In 1979, during the annual IAI conference, a resolution was approved prohibiting members from giving testimony on qualified identifications (likely, possible, probable, etc.). An identification, according to the IAI, can only be negative, impossible or certain. This resolution was widely debated before being revised and accepted in 1980. The IAI experts then rejected the idea of using dactyloscopic proof as corroborative evidence in the following terms (Resolution VII Amended or Resolution V): '[The delegates of the IAI] state unanimously that friction ridge identifications are positive, and officially oppose

any testimony or reporting of possible, probable or likely friction ridge identifications found on hands and feet, [...]'

This approach means that in countries using numerical standards, for instance Great Britain, a fingerprint comparison with 16 concordant minutiae is an identification, whereas a comparison presenting only 15 points is not probative evidence and should not be presented in court. Such a rule can easily result in difficult situations when, for example, a suspect arrested and identified in one country on the basis of an identification with 12 points is to be judged in a neighbouring country where 16 points are explicitly required.

The present author considers that the resolution voted by the IAI in 1980 prohibiting this form of testimony is opposed to the scientific principles governing the interpretation of transfer traces in criminalistics. It is erroneous to consider that dactyloscopic evidence is only a dichotomic form of proof, uniquely applicable to identifications or nonidentifications. The interpretation cannot be so drastic, so clear-cut; there is an increasing scale from exclusion to identification. In fact, there is no logical reason to suppress gray levels between white (exclusion) and black (identification). Each piece of evidence is relevant if it tends to make the matter at issue more or less probable than otherwise. Dactyloscopic evidence can be combined with other evidence (forensic or not) and fit into the particular context of each case. The refusal of qualified opinions in dactyloscopy is a policy decision (even if the distinction of the arguments (policy or scientific argumentation) is not so clear in the literature).

Often, probabilities have been excluded from dactyloscopy arguing that each portion of a fingerprint, no matter how small, can only have come from one finger. The individuality of a particular papillary arrangement, and even the individuality of the process and the contours of the ridges, is without question. However, the transfer of this information necessarily implies a loss of detail – a loss of information – that reduces this individuality with respect to the trace evidence itself. It is the individuality of the papillary impression that is at question and not the individuality of the skin surface that produced it. The transfer of material in criminalistics, as defined by Locard, logically implies, by definition, a loss of information. In dactyloscopy, this loss may be of two types: quantitative (due to the limited size of the trace) and qualitative (blurring, bad definition, loss of pore details, etc.). The concept encapsulates a continuum of values in that a feature may have an identifying value ranging from very low to very high. Locard's third directive is fundamental and permits dactyloscopic proof to be on the same level as other types of transfer traces. Some will no doubt argue that judges expect dactyloscopy to be univocal and without compromise. This idea is false. It is obvious that any judge will prefer indisputable forms of proof, but would also wish to be informed of any evidence that can lead closer to the truth. Fingermarks have already been considered by jurors from a probabilistic point of view. By their admission, these traces may be classed with all other forms of evidential material in the judicial system. This issue has been addressed by New Zealand when the Court of Appeal not only ruled that there are no numerical standards for an identification but has also stated that:

- much useful evidence was being kept out of the courts on decision of technicians;
- fingerprint evidence should (of course) be combined with other evidence to determine whether or not the accused was present.

The absence of extensive statistical data on fingerprint individuality can be viewed as the main reason to prevent giving qualified opinions. Statistical data could make dactyloscopists feel more comfortable in this area and make statements based on sound data obtained from significant samples.

Research results only describe one factor contributing to fingerprint individuality, which is the probability of the configuration of minutiae (basic forms and combinations) on a surface $P(C)$. A computer program has been developed to enable the acquisition of data on an extensive sample of nearly a thousand fingerprints selected from more than 100 000 dactyloscopic forms. Nine types of minutiae (ridge ending, bifurcation, island, double bifurcation, hook, lake, opposed bifurcation, etc.) were considered, taking into account their orientation and length when defined.

A model giving an upper bound estimate $P(C^*)$ of the probability $P(C)$ has been proposed and successfully tested. To illustrate the impact of these results, let us consider two different configurations of the same size taken from the periphery region of two ulnar loops (**Fig. 1**).

The probabilities $P(C1^*)$ and $P(C2^*)$ have been calculated and are as follows:

$$P(C1^*) \cong \frac{1}{40000}$$

$$P(C2^*) \cong \frac{1}{1.42 \times 10^9}$$

These probability figures strongly suggest the possibility of evaluating dactyloscopic evidence in a probabilistic way. When compound minutiae are

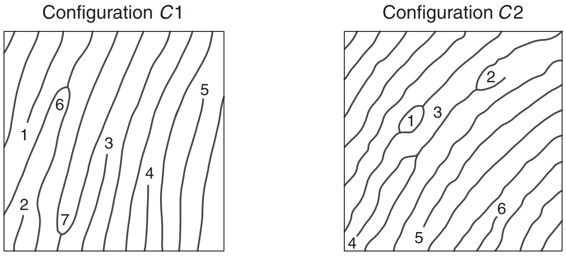


Figure 1 The configuration *C*1 shows a skeletonized fingerprint pattern with two bifurcations and five ridge endings. The configuration *C*2 shows a pattern with one lake, two bifurcations facing each other, a hook and three ridge endings.

Table 3 Results of the proficiency tests made through the CTS program from 1995 to 1998

Year	*Answer sheets*	*Comparisons per sheet*	*Erroneous identifications*
1995	158	7	48 (by 34 laboratories)
1996	191	11	8 (by 6 laboratories)
1997	204	11	16 (by 13 laboratories)
1998	219	11	21 (by 14 laboratories)

observed on a surface, the calculation possibilities are extended and lead to probabilities which are of impressive value as evidence. It is possible to obtain, with some specific minutiae on a surface (or in the absence of minutiae on a large papillary surface), probability figures which exceed the value proposed in other forensic fields (e.g. DNA), even with less than 12 points.

The Move Towards Quality

Proficiency tests results

An indicator of quality may be searched in the results of the proficiency testing in the profession. The Collaborative Testing Service, Inc. (Herndon, VA), now in association with the IAI proposes declared annual proficiency tests to agencies around the world. A summary of the results for the latent prints examination is given in **Table 3**.

These results clearly indicate that error is possible and that some examiners do not reach adequate quality standards. But the evaluation of the standard of proof (or the quality of the evidence) cannot be limited to the examination of proficiency testing programs. Indeed, the corrective actions taken in an agency, following an erroneous identification, are far better indicators of quality than the report of errors (or successes). A global assessment of quality is then needed.

Total quality management

A total quality management system, based for example on the European Foundation for Quality Management (EFQM) model, can be recommended and is the basis of the most recent changes in the fingerprint practice. The various aspects of such a system are illustrated in **Fig. 2**, the program is divided here into three parts: fingerprint examiners, processes and products.

For each category the quality management scheme must specify the needs and objectives, along with the

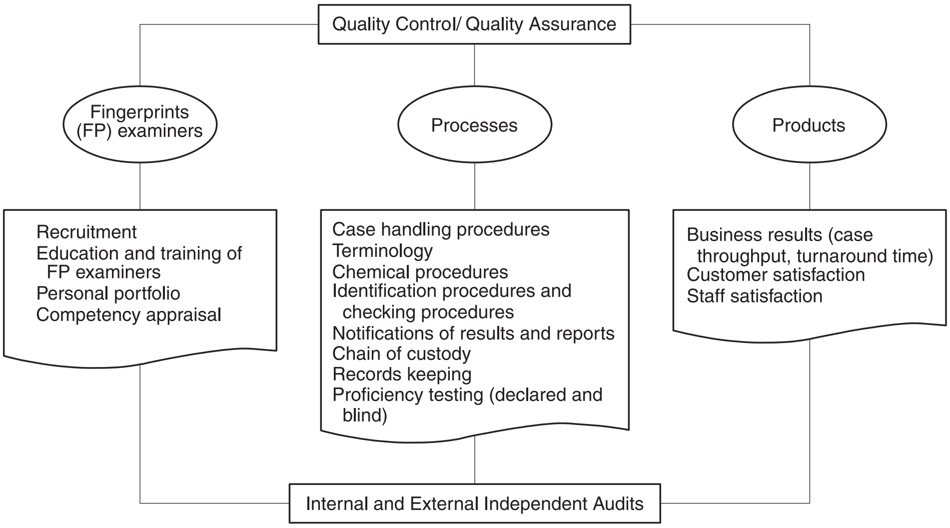


Figure 2 Outline of a total quality management system for a fingerprint service.

quality indicators used to monitor the system and the corrective actions available. Audits (internal and independent) permit the claims to be checked regularly. In such a general context, proficiency testing must be regarded as only one block of the building. The value of a fingerprint identification will be improved by a strict quality assurance program.

Conclusion

It is obvious that Locard proposed the basic concepts determining the value of fingerprint evidence. Locard's opinions were enlightening and, above all, the most topical.

With regard to numerical standards, there is no scientific argument in favor of any predefined number of minutiae (in the absence of discordance) in order to pronounce an identification. Every comparison bears its own specificity which refrains from defining *a priori* the amount or volume of information needed for the identification. It forces us to rely on experience, professionalism and integrity of fingerprint examiners. Even if some practitioners rely on minimum numbers, they do not escape, thanks to the arbitrary limit, from the quality issues because, depending on the examiners, the interpretation of the number of minutiae in a comparison may vary to a large extent.

Finally, following this description of dactyloscopic practices, what is the standard of proof?

Definitely, the standard is not a predefined number of minutiae. Indeed, insisting dogmatically on any particular minimum number of points is not an effective way of insuring quality. The way forward is the adoption of a scheme of total quality management dealing with the various aspects of education and training, laboratory procedures and audits.

Concerning the possibility of giving qualified opinions based on dactyloscopic evidence (likely, possible, probable, etc.), it is clear to the author that from a scientific point of view, there is no argument to justify the so-called 'positivity' of the field. Nothing opposes the use of fingerprint evidence as corroborative evidence in the same way as other transfer traces.

See also: **Fingerprints (Dactyloscopy)**: Visualization; Identification and Classification.

Further Reading

Asbaugh DR (1999) *Quantitative–Qualitative Friction Ridge Analysis*. Boca Raton: CRC Press.

Balthazard V (1911) De l'identification par les empreintes digitales. *Comptes rendus des séances de l'Académie des Sciences* **152**:1862–1864.

Champod C (1995) Locard, numerical standards and 'probable' identification. *Journal of Forensic Identification* **45**:132–159.

Champod C (1996) *Reconnaissance Automatique et Analyse Statistique des Minuties sur les Empreintes Digitales*. Thèse de doctorat. Lausanne: Institut de Police Scientifique et de Criminologie. Lausanne, Université de Lausanne.

Evett IW and Williams R (1996) A review of the sixteen point fingerprint standard in England and Wales. *Journal of Forensic Identification* **46**:49–73.

Lambourne GTC (1984) *The Fingerprint Story*. London: Harrap.

Locard E (1914) La preuve judiciaire par les empreintes digitales. *Archives d'Anthropologie Criminelle, de Médecine Légale et de Psychologie Normale et Pathologique* **29**:321–348.

Robertson B and Vignaux GA (1995) *Interpreting Evidence – Evaluating Forensic Science in the Courtroom*. Chichester: John Wiley.

Stoney DA (1985) *Quantitative Assessment of Fingerprint Individuality*. PhD Dissertation. Graduate Division of the University of California. Berkeley, University of California.

Wertheim PA (1996) The ability equation. *Journal of Forensic Identification* **46**:149–159.

Visualization

J Almog, Casali Institute of Applied Chemistry, The Hebrew University, Jerusalem, Israel

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0518

Introduction

Law enforcement officers and the general public associate fingerprints, more than any other category of physical evidence, with criminal investigations and apprehension of criminals. Fingerprints definitely prove the presence at a crime scene of the individual who made the impressions, and they establish that said individual once had contact with a given item now constituting evidence.

Each ridge of the fingers and palms, and the soles of the feet bears a row of sweat pores which constantly exude perspiration (**Fig. 1**). In addition, the ridges of the fingers and the palms are in intermittent contact with other parts of the body, such as hair and the face, and with various objects, which may leave a film of grease or moisture on the ridges. When an object is touched, an outline of the ridges of the fingers or palm is deposited. This outline is called latent impressions, or latent fingerprints.

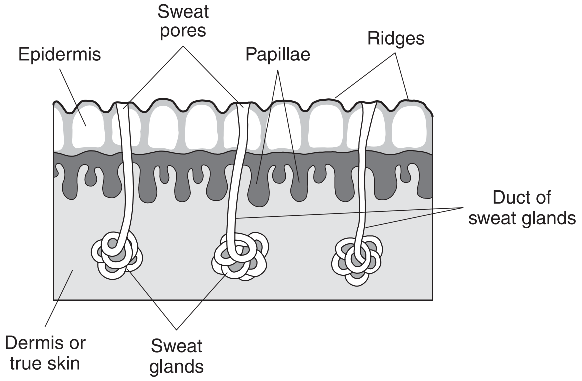


Figure 1 Cross section of friction skin.

Latent prints are, perhaps, the most prominent example of Locard's exchange principle: 'Every contact leaves traces'; hence, the great importance of visualizing traces and transferring them onto useful and valuable evidence.

The detection and imaging of latent fingerprints is unique in forensic science. The objective is not to detect and quantify specific chemical substances *per se*, but rather to map the distribution of these chemicals on various surfaces. Indeed, one of the most exciting and dynamic areas of research in forensic science today is the application of physicochemical techniques to the visualization of latent fingerprints. Changes are occurring very rapidly as researchers uncover a variety of processes applicable to the visualization of latent prints. For many years progress in this field was minimal, as fingerprint specialists traditionally relied on three chemical techniques – iodine, ninhydrin and silver nitrate – to reveal the presence of fingerprints. Many more techniques have been developed over the last two decades, and special attention has been given to techniques based on fluorescence, or fluorogenic processes. In designing improved fingerprint reagents, it is essential to have a sound knowledge of the chemical composition of both fingerprint deposits and of the surfaces on which they are found. Much work in this aspect was conducted by the Police Scientific Development Branch (UK) in the 1970s.

To achieve successful results, it is not necessary for a fingerprint technician to know the exact chemical composition of a latent print residue or the chemical processes involved. In the majority of instances, they must see the surface to be examined, and then decide on the proper sequence in applying chemical techniques.

Not all attempts to recover latent fingerprints are successful, even with proper handling and application of the appropriate fingerprint techniques. Failure to recover latent fingerprints does not, in itself, necessarily mean that there were no prints on the item or that they were wiped off. The inability to recover latent prints can be due to numerous factors. It is a function of perspiration of the individual leaving the latent mark, surface type, environmental conditions and the age of the print.

Chemical Composition of Fingerprint Deposits

A latent print is a complex mixture of natural secretions and environmental contaminants. Eccrinic, apocrinic and sebaceous glands are responsible for natural secretions of the skin. Of these, only eccrinic glands are found in ridged skin, where they are abundant. Studies have shown that there are 550–950 sweat pores per square centimeter in finger ridges, and less (400) in the palms and soles.

The apocrine glands are found in the armpits, chest, abdominal and genital areas, whereas sebaceous glands are found on the forehead, around the nose, on the back and genital areas.

As a result, eccrinic gland secretions are present to some degree in every latent fingerprint; contamination by sebaceous secretions (sebum) is also very common from people touching their faces.

Approximately 99.0–99.5% of perspiration is water. The remaining residue consists of small amounts of organic and inorganic materials. The salts predominant in perspiration are sodium and potassium chlorides, with the organic fraction containing mainly amino acids, urea and lactic acid. Free fatty acids, triglycerides and wax esters prevail in sebaceous secretions (**Tables 1–3**).

Table 1 The main constituents of the secretion of the sweat glands (excluding water)

Gland type	*Inorganic constituents*	*Organic constituents*
Eccrine	Chlorides	Amino acids
	Sodium	Urea
	Potassium	Lactic acid
	Ammonia	Sugars
	Sulfates	Creatinine
	Phosphates	Choline
		Uric acid
Apocrine	Iron	Proteins
		Carbohydrates
		Cholesterol
Sebaceous	–	Fatty acids
		Glycerides
		Hydrocarbons
		Alcohols

From Knowles A M (1974) Development of latent fingerprints on paper and fabrics, *Proceedings of the Conference on The Science of Fingerprints*. Police Scientific Development Branch. London: Home Office.

Table 2 Levels of amino acids, chlorides and urea in fingerprint deposits

Compound	*Levels in fingerprints* ($\mu g\ cm^{-2}$)	*Levels in 120 types of paper* ($\mu g\ cm^{-2}$)
Amino acids (as serine)	Range 0.120–0.720 Mean = 0.250	0.004–1.000
Chlorides	Range 0.270–1.500 Mean = 0.670	0.500–6.000
Urea	Range 0.090–0.720 Mean = 0.250	No detectable levels in the papers tested in this study. Some surface-coated papers do have high levels of urea.

From Knowles AM (1974) Development of latent fingerprints on paper and fabrics. *Proceedings of the Conference on The Science of Fingerprints*, Police Scientific Development Branch. London: Home Office.

Table 3 Main lipid classes in sebum

Component	%
Free fatty acids	30
Triglycerides	27
Diglycerides	4
Monoglycerides	2
Wax esters	22
Squalene	10
Cholesterol	2
Cholesterol esters	2
Hydrocarbons	1

From Knowles AM (1974) Development of latent fingerprints on paper and fabrics. *Proceedings of the Conference on The Science of Fingerprints*, Police Scientific Development Branch. London: Home Office.

In addition to natural secretions, latent fingerprints may often be contaminated by other materials present in the environment and picked up on the skin. It is usually impossible to determine the major constituents from simple visual examination, hence, apart from when the marks are obviously made in a contaminant such as blood, ink, dust or grease, the latent prints are considered to contain mainly sweat.

The possibility of visualizing a latent fingerprint depends on the constituents of the original fingerprint (the individual person), the nature of the surface, the time elapsed since deposition and the storage conditions. Factors such as temperature, humidity and exposure to light and dust have a strong adverse effect on the durability of latent prints.

Latent Fingerprint Techniques: Development of Latent Fingerprints

A latent fingerprint technique is any method, process or procedure used to visualize or enhance latent fingerprint impressions for preservation and identification.

Despite the wide range of surfaces and detection methods described in the literature, even the most advanced laboratories do not usually employ more than ten different methods that are carried out in various combinations and modifications. These are: visual examination, dusting with powders, cyanoacrylate fuming, vacuum-metal deposition and, occasionally, iodine, for smooth, nonabsorbent surfaces; and visual examination, diazafluorenone (DFO), ninhydrin, physical developer and, occasionally, iodine for porous surfaces, such as paper and cardboard. Additional methods used less frequently are silver nitrate (for rough surfaces such as wood), small particle reagent (SPR) (for wet, smooth or dusty surfaces), and 'secondary' methods such as contrast enhancement by fluorescent powders, or intensely colored substances such as gentian violet. Weak prints in blood can be enhanced by chemical reagents, for example amido black, luminol or tetramethyl benzidine. Since the person conducting the test will normally seek the optimal development method for the surface to be examined, the range of tests is divided according to surfaces and the best test for each.

Only rarely will just one test detect all the latent prints on a surface, though sometimes there is no practical way to employ multiple methods (cost-effectiveness considerations or pressure of time). At present, at least in the investigation of serious crimes, forensic science laboratories will employ a battery of detection methods on each surface. The methods are used in sequence and after each stage, should prints develop, the results are photographed and the process continued.

Obviously, the process commences with the nondestructive methods followed by methods that are only partly destructive, and concludes with techniques that leave permanent traces. Another consideration for the sequence of methods chosen is that a prior method should not interfere with a later method.

General, nondestructive techniques: visual examination by optical means

Optical detection has the advantage of not destroying surfaces and latent fingerprint deposits. As a result, these techniques do not preclude the later application

of other fingerprint techniques. Observation under white light may reveal latent fingerprints that can be photographed without any further treatment. More complex optical methods may disclose prints that are otherwise invisible.

White light detection White light, particularly in strong illumination, is one of the most useful tools in fingerprint visualization. Strong white light illumination is used in nearly every case in which fingerprint evidence is sought. Proper alignment between the light source and field of vision is normally determined by trial and error. A print can be invisible with the light in one position, but may be most apparent with the same light in another position.

Episcopic coaxial illumination Highly reflective surfaces, such as glass, polished metals and certain plastics, pose a problem in fingerprint detection under white light due to reflection. In such cases, latent prints can often be detected using episcopic coaxial illumination. This is a special illumination technique which uses a semitransparent mirror to observe the reflection of light perpendicular to the surface (**Fig. 2**). The light is diffused by the fingerprint deposit, but specularly reflected by the surface. The print is therefore visible as dark ridges against a light background. This technique also gives excellent results after cyanoacrylate treatment.

Ultraviolet illumination Ultraviolet illumination as a nondestructive detection method for latent fingerprints is far less common, since natural fingerprint deposits do not show any particular response in this domain. Certain environmental contaminants, however, such as greases or food ingredients do fluoresce under long-wave UV illumination.

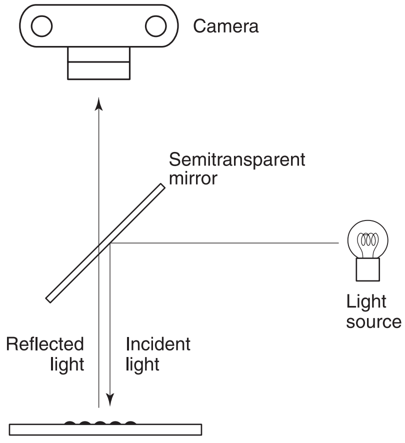


Figure 2 A diagram for fingerprint detection by episcopic coaxial illumination. (Margot P, Lennard C (1994) *Fingerprint Detection Techniques*. Switzerland: Université de Lausanne.)

Due to the fact that today UV light sources are inexpensive, easily operated and readily available, their use (long wavelength, 354 nm) in many laboratories is a common, nondestructive procedure after examination in white light.

Detection systems that involve short-wave UV illumination (< 300 nm) have been described in the literature. Detection is based on either reflection of the fingerprint deposit, which differs from that of the background, or on the induced short-wave UV fluorescence of certain sweat components. The detection system involves a strong source of short-wave UV light (frequency quadrupled Nd:YAG lasers or mercury lamps) and UV-sensitive CCD camera equipped with quartz lenses. Although there are some commercially available systems that use short-wave UV illumination for fingerprint detection, these are still considered experimental. In addition, they are not totally nondestructive, since prolonged illumination by short-wave UV light may irreversibly alter DNA from the exhibits. For the same reason, special safety precautions must be employed for short-wave UV work, such as the wearing of safety goggles and skin protection gear.

Lasers and alternate light sources A most important development in the field of latent fingerprint detection was the discovery in 1976 that latent fingerprints on various surfaces could be revealed by their inherent luminescence when illuminated with an argon-ion laser.

It was later found that in actual casework only a small fraction of untreated prints could be visualized by this technique. Nevertheless, if an argon-ion laser is used, it should always precede more destructive methods. The detection is based on the presence of luminescent compounds, such as riboflavin, in the sweat components, but it is more likely to obtain good results when the fingers are contaminated with foreign luminescent substances. An outstanding example of the value of this nondestructive technique was the detection, by the FBI Laboratory in 1984, of the 42-year-old thumbprint of the war criminal, Valerian Trifa, on a postcard sent to Heinrich Himmler in 1942. The procedure used to detect latent prints with the laser is relatively simple, and evidence is not altered. Latent prints have been detected by their inherent luminescence on a wide variety of surfaces, including metals (e.g. firearms), paper, adhesive tapes, plastic and polystyrene foam. Positive results have been reported even on human skin.

The light beam emitted from the argon-ion laser (green and blue–green in color, the main two lines, 488 and 514.5 nm) is dispersed by a fiberoptic cable to cover an area of approximately 10 cm in diameter.

The illuminated area is viewed through a long-wavelength-pass filter that transmits the fingerprint fluorescence but blocks the reflected laser light. Observed fingerprints are photographed through the same filter (**Fig. 3**).

Alternate light sources for fingerprint detection began to appear in the mid 1980s. They were smaller, lighter, less expensive and more diverse than the lasers. In 1983, in Australia, a modified xenon arc lamp was developed. At about the same time, the 'Quaser' light source was developed by the Scientific Research and Development Branch of the UK Home Office, and the 'Lumaprint' lamp was designed at the National Research Council of Canada. In many forensic laboratories alternate light sources (also referred to as 'forensic lights') have become the main nondestructive tool for the search of latent fingerprints.

These devices are, basically, powerful sources of white light equipped with an appropriate set of filters, which transfer only the desired fraction of light that hits the material under examination and excites fingerprint or background fluorescence. Although their performance in fingerprint visualization is generally comparable with that of lasers, it was reported by groups in Israel and Canada that, in actual casework, latent prints were encountered that could not be visualized by the alternative light sources whereas the argon-ion laser or the copper-vapor laser did detect the prints. However, the main contribution of lasers and their substitutes to fingerprint visualization is not in the detection of inherent fingerprint luminescence, but in the excitation of luminescence after the latent prints have been chemically treated by fluorescent or fluorogenic reagents.

Visualization techniques for smooth, nonporous surfaces

Surfaces such as glass, paint, plastic and metals are included in this category.

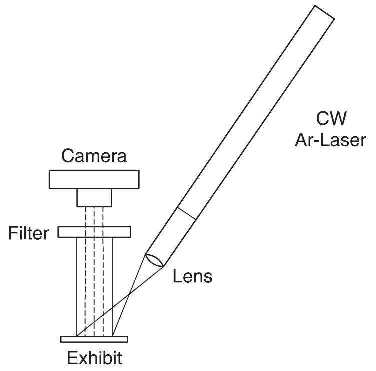


Figure 3 A diagram for fingerprint search by argon-ion laser. (From Menzel ER (1980) *Fingerprint Detection with Lasers*. New York: Marcel Dekker.)

Powder dusting The simplest and most commonly used procedure for latent fingerprint development is powder dusting. Visualization is based on the physical adherence of fingerprint powders to the moisture and oily components of skin ridge deposits, but the exact mechanism is not fully understood. Application of powder to latent prints by brushing is a simple technique and yields instantly apparent prints. Only minimal training is required to obtain satisfactory results. Powder dusting, however, is considered an insensitive technique and only relatively fresh latent prints are normally developed. Hundreds of fingerprint powder formulae have been developed over the years. Aluminum powder is composed of microscopic aluminum flakes coated with stearic acid. Fluorescent powders luminesce under various wavelength illumination and may be used on reflective surfaces. Magnetic powders are composed of iron flakes, mixed with copper or aluminum flakes; they are applied by a magnetic wand which allegedly avoids brushing and the destruction of fragile prints. Choice of the appropriate powder and application method is often made according to experience and personal preference. Fingerprint brushes are made of hairs, feathers or fiberglass. Fiberglass brushes last longer than the other two types, and have gained wide acceptance in recent years.

Cyanoacrylate fuming In 1977, the National Police Agency of Japan made an amazing discovery that had an immense impact on the art of latent fingerprint detection. They noticed that fumes of the high-strength quick glue, 'Superglue', developed latent fingerprints by selectively polymerizing on the ridges in the form of a hard, white crust. This technique was quickly adopted by fingerprint practitioners the world over and also received much attention from researchers who optimized its performance and extended its range of application.

'Superglue' is composed basically of a cyanoacrylate ester, normally methyl or ethyl cyanoacrylate. These are colorless liquids with relatively high vapor pressure. Cyanoacrylate fumes polymerize upon reaction with latent fingerprint deposits. The polymerization is thought to be catalyzed by the water and possibly other perspiration constituents (for mechanism, see **Fig. 4**).

The use of cyanoacrylate vapor has greatly enhanced the possibility of developing latent prints on most nonporous surfaces, especially plastic bags and other pliable plastics, but also on styrofoam, carbon paper, aluminum foil, finished and unfinished wood, rubber, metals and even smooth rocks. The contrast of developed fingerprints may be enhanced by the application of dyes and powders.

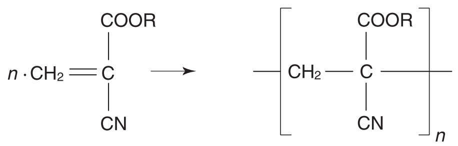


Figure 4 The cyanoacrylate process.

Equipment required for cyanoacrylate fuming includes a fuming tank, cabinet or other suitable container equipped with a proper ventilation system. The samples to be treated are suspended in the tank, and a few drops of liquid cyanoacrylate are dripped into a container at the bottom. A container of water is also placed in the tank to insure sufficient humidity for the development process (prints of poor contrast are obtained if the humidity is too low). Variations on this procedure involve applying heat or chemicals to accelerate the vaporization; varying exposure times for optimal results; controlling the generation of fumes under low pressure ('vacuum cyanoacrylate'); using portable field devices to fume large objects; staining techniques for improved contrast; the use of the polymer, polycyanoacrylate, as the source of the fumes; incorporation of cyanoacrylate into a gel matrix (Superglue 'pouch'); using various alkyl groups in the ester (ethyl ester is the most common); and removal of excess polycyanoacrylate from prints overexposed to Superglue.

Vacuum metal deposition (VMD) Vacuum metal deposition (VMD) utilizes vacuum coating technology for the evaporation of metals and the deposition of thin metal films. It is based on a phenomenon that has been a great nuisance in industrial growing of thin layers of metals on various surfaces, where fingerprint contamination hinders the deposition of the metallic film following metal evaporation under vacuum. The potential of this method for fingerprint detection was first reported in 1968, and thoroughly investigated a few years later but practical work started in 1976.

VMD is considered an extremely sensitive and useful technique for fingerprint detection on a variety of smooth, nonporous surfaces, particularly plastic packaging materials such as polythene, leather, photographic negatives and prints, plastic moldings and glass.

Gold is evaporated under vacuum to form a very thin, invisible layer on the surface under examination. A second layer, this time, zinc, is deposited in the same manner. The zinc preferentially coats the regions between and around the ridges, where the gold is more exposed. Thus, the ridges appear transparent whereas the furrows and the background are dark (since they are plated with a zinc layer). Interestingly, the original second metal that gave excellent results was cadmium but, due to its toxicity, it was replaced by zinc.

VMD can sometimes reveal fingerprint detail where other techniques have failed. Good results can also be obtained after cyanoacrylate development.

One of the advantages of VMD is extreme sensitivity, even in cases of old fingerprints and prints which have been exposed to water. Disadvantages are the initial cost of high vacuum equipment, the need for expert operators, and the fact that it is a most time-consuming process.

Iodine fuming Iodine fuming is one of the oldest methods used for the chemical development of latent fingerprints. Its use was mentioned as early as 1892. It can develop prints quickly and neatly on porous and nonporous surfaces.

Iodine vapors are absorbed by fingerprint residue to form a brown image which fades rapidly. For many years, it was assumed that iodine adds to the double bonds of the unsaturated fatty acids present in perspiration, but there is convincing evidence that this is a physical, rather than a chemical, reaction, and that iodine fumes are also absorbed by other sweat components, such as saturated fatty acids and even water. The technique is simple but not very sensitive. It reveals latent prints not more than a few days old. The developed prints must be photographed immediately before they fade. Alternatively, they can be fixed by chemical means such as benzoflavone or tetrabase. Although the iodine fuming method has been ignored somewhat due to more modern, faster techniques such as lasers and cyanoacrylate fuming, it remains an important method due to its simplicity, speed and low cost. Disadvantages are low sensitivity to old prints and the toxic and corrosive nature of the fumes.

Iodine fumes are obtained by slightly warming iodine crystals. Control of the fumes is achieved by using the crystals in an iodine gun or fuming cabinet. Iodine blowing guns are made of a glass or hard plastic tube. A drying agent such as calcium chloride is used to dry the vapors. The iodine crystals sublime by the heat of the breath, which is augmented by the warmth of the hand which is cupped around the tube containing the crystals. The vapor is blown onto the specimen. Fuming cabinets are normally made of glass. The fumes are generated by placing a small heating device under an evaporating dish containing the iodine crystals. The specimens are then suspended above the dish.

Iodine solution method modification of the fuming technique, that uses iodine in solution, is used by law-enforcement agencies in the UK. A solution of iodine in cyclohexane containing 7,8-benzoflavone is used at

crime scenes to search for latent fingerprints on surfaces such as wallpaper, emulsion-painted walls and aged glass-painted surfaces. The reagent is applied by spraying or brushing, and the fingerprints show up as dark-blue ridges. This technique is particularly effective for revealing fresh marks.

Visualization techniques for porous surfaces

The main surfaces in this category are paper and cardboard. The search for latent prints must start, as for nonporous articles, with nondestructive visual examinations and continue with chemical techniques that can be augmented by appropriate light sources.

Ninhydrin Perhaps the most meaningful breakthrough in the chemical detection of latent fingerprints on paper occurred in 1954, when ninhydrin (**Fig. 5**) was proposed for the development of latent fingerprints. Since then ninhydrin has become the most consistently successful reagent for developing latent fingerprints on documents and other porous surfaces. Ninhydrin reacts with amino acids to produce a highly colored compound known as Ruhemann's purple (named after Ruhemann, the scientist who discovered and interpreted this reaction in England, in 1910).

Amino acids on paper are relatively stable and do not migrate with age. Ninhydrin is, therefore, a most suitable reagent to reveal even old prints on paper. Its use is simple, but skill and training are required to obtain good results.

In 1955, a formulation for fingerprint development was proposed which consisted of acetone as a solvent and acetic acid to enhance sensitivity. Since then, numerous formulations and development conditions have been investigated by various researchers. Among the parameters studied were: solvents, concentrations, temperatures and heating times, pH, modes of application and humidity conditions. Currently, the most common formulation is a modification based on the nonpolar solvent 1,1,2-trifluorotrichloroethane (fluorisol, freon 113). The use of freon 113 virtually eliminates the problems of ink running and is also nonflammable. This formulation is known by the abbreviated process name NFN (nonflammable ninhydrin). Due to its adverse effect on the ozone layer,

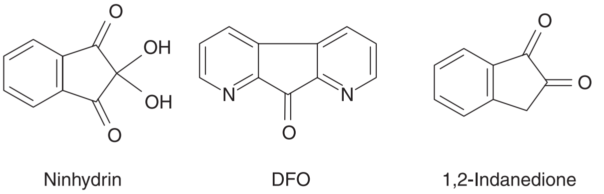


Figure 5 Formulae for ninhydrin; 1,8-diaza-9-fluorenone (DFO) and 1,2-indanedione.

freon 113 is now banned in many countries and alternative solvents are currently being studied.

Generally, items being examined are dipped into ninhydrin solution for a few seconds, dried in air, then warmed in an oven at temperatures not exceeding 80°C and c. 65% relative humidity. Fingerprint marks appear as purple impressions. The reagent may also be applied by spraying or brushing. Weak ninhydrin prints can be improved by converting them into fluorescent impressions by treatment with zinc chloride. This process, originally suggested in 1982, and further study suggested the use of cadmium instead of zinc. The structure of the fluorescent complex was clarified in 1987.

DFO (1,8-diaza-9-fluorenone) DFO is one of the most recent and significant developments in reagents for the chemical development of latent fingerprints. Its introduction into practical work has brought about a considerable increase in the number of latent fingerprints that can be revealed on paper items. Its specific reactivity was originally discovered by Grigg (University of Belfast) and Pounds (CRE, British Home Office), during a project on ninhydrin analogues. Among other compounds, the heterocyclic ketone, 1,8-diaza-9-fluorenone (DFO, Fig. 5) was studied and its solution was found to produce strong luminescent impressions of latent fingerprints on paper. A mechanistic explanation to the fluorogenic process was also suggested, based on a reaction between DFO and amino acids.

Some advantages of the DFO reagent are high sensitivity, relatively broad excitation and emission bands, adequate solubility in nonpolar solvents and the possibility of using ninhydrin and physical-developer after the application of DFO. In addition, there is no need for secondary treatment with zinc to excite the luminescence. To develop latent fingerprints, paper and cardboard items are dipped in a dilute solution of DFO in freon 113 and acetic acid. The articles are heated in the oven and illuminated by an argon-ion laser or one of the alternate light sources emitting in the blue, blue–green or green domain. Observation of the luminescent impressions is through a cut-off filter that transmits light above 520 nm. In some cases, the developed prints may also have a pale purple color, which is much weaker than that of ninhydrin-developed prints, but may indicate the presence of latent fingerprints on the examined article.

Physical developer (PD) Physical development is a photographic process based on the formation of gray silver deposits from an aqueous solution containing silver ions in the presence of ferrous/ferric redox couple and a detergent. A review of the early work

on techniques using stabilized physical developer for fingerprint development found that metallic silver is deposited on the lipid material present in the fingerprint residue. The detergent prevents premature precipitation. It was later found that the process can also be used after treatment with DFO or ninhydrin. It can produce a significant number of extra prints, since PD reacts with different constituents of the fingerprint deposit. It is a very effective reagent for developing latent fingerprints on wet paper. Other articles that can be processed by PD are adhesive strips on envelopes, adhesive tapes, the emulsion side of photographs, and certain types of currency notes. PD reagent is delicate to prepare, since the solution is not very stable. Also, silver sometimes precipitates on the entire surface of the paper. Hence, PD should be applied at the end of any detection sequence on paper. Its successful application requires considerable experience. The working solution must be prepared carefully in clean glassware and only good quality distilled or deionized water should be used. The detergent contains *N*-dodecylamine acetate and synperonic NP8; silver nitrate is the source of silver ions, and the redox couple is a mixture of ferric nitrate and ferrous ammonium sulfate in dilute citric acid. The sample must be washed in distilled water to remove surface contamination before it can be treated with the reagent. Certain paper items require a prewash with maleic acid to reduce background discoloration. The development process occurs by soaking the articles in the reagent, rinsing with distilled water, and drying.

Modified PD Several modified PD procedures have been reported: the use of scanning electron microscopy for observation (to reduce background interference); conversion of the silver deposit to radioactive silver sulfide and recording the image by autoradiography; and particularly, Saunders' and Cantu's 'multimetal deposition' in which colloidal gold provides a nucleation site for the silver precipitation.

Silver nitrate One of the oldest methods for chemical development of latent fingerprints, silver nitrate treatment, has been in use since 1891. Silver ions react with the chlorides present in eccrine perspiration. The colorless silver chloride thus formed turns black rapidly on exposure to light. The procedure is simple but it may produce a high background reaction which obscures fingerprints. In light of more advanced chemical methods such as DFO, ninhydrin and physical developer, the silver nitrate technique is considered obsolete, at least for paper items, by many forensic science laboratories. It may still be of some value on raw wooden surfaces.

Small particle reagent (SPR) Small particle reagent, also known as 'suspended particle reagent' (SPR) consists of a suspension of fine molybdenum disulfide particles in detergent solution. The particles adhere to the fatty constituents of latent print residues to form a gray molybdenum disulfide deposit. The method was patented in 1979. Its great advantage is that it can develop latent prints on nonporous, dusty or wet surfaces, where the use of powders is excluded. It can be applied in a dish or by spraying. The latter technique is less sensitive and should be considered only when dish application is impossible. White SPR, based on fine zinc carbonate particles, can be used on dark surfaces, as well as fluorescent SPR.

Other methods Additional special techniques have been designed for fingerprint development on problematic surfaces. For example, the *radioactive sulfur dioxide* method was developed to visualize latent fingerprints on fabrics and adhesive tapes. Its application requires special facilities and trained staff, hence, it is used only in high-profile cases.

The National Police Agency of Japan uses *ninhydrin hemi-acetal*, a nonpolar derivative of ninhydrin in petrol–ether solution, to develop latent prints on 'thermal' paper. The US Secret Service and the Metropolitan Police Forensic Science Laboratory (London) advocate subliming *p-dimethylaminocinnamaldehyde* (*p*-DMAC) for the same purpose. A suspension of *Sudan black* can be used to develop sebaceous components of latent prints on surfaces contaminated with grease or foodstuffs. Many more variations are in routine use by fingerprint practitioners. Some latent fingerprint techniques mentioned in the professional literature, but rarely used, are listed below. *Osmium tetroxide* (which is extremely toxic) reacts with unsaturated fatty acids, to deposit, on the ridges, dark osmic acid impressions on both porous and nonporous surfaces. *Ruthenium tetroxide* reacts in a fashion similar to osmium tetroxide. It is less toxic than osmium tetroxide but is, nevertheless, an eye and respiratory tract irritant. *p-Dimethylaminocinnamaldehyde* (*p*-DMAC) reacts with urea present in latent print deposits on paper, to form red impressions. Due to the migration of urea from the deposits, it only reacts well with fresh prints. (p-DMAC is now considered to be a much more efficient amino acid fluorogenic reagent in the vapor phase.)

Heat has been mentioned by several experts as a latent fingerprint technique for documents. The organic substances in the fingerprint residue are charred and appear as dark impressions. Although heat cannot be considered a practical latent fingerprint method, a number of arson cases have been solved as a result of latent prints developed by heat.

Such prints can be of paramount importance when processing of the evidence by regular techniques is precluded due to excessive charring. The *flame* technique, for softening dried impressions and producing soot that adheres to the latent marks, was used successfully in several cases in the past. Its use is quite scarce, but again, in arson cases, latent prints may be developed by soot.

Particular cases

Enhancement of bloody fingerprints Special techniques are often required for the enhancement of weak fingerprints in blood. A nondestructive optical technique has been suggested based on the reflection of fresh blood upon illumination at 400 nm (violet). The bloody prints appear as white impressions on a darker background. Ultraviolet fluorescence can also provide useful enhancement by improving the contrast on dark surfaces.

Two groups of reagents for chemical enhancement of bloody prints are used. The first contains colorless substances catalytically oxidized in the presence of hemoglobin, to form colored products. *Tetramethylbenzidine* is the most common reagent in this category. It replaced the carcinogenic *benzidine*, the use of which was banned in the US in the early 1970s. Other substances belonging to 'heme-reagents' are *o*-tolidine, phenolphthalin and leukomalachite green. The second group contains reagents that stain proteins. Their solutions react with proteins to form colored complexes. *Amido black* is a typical reagent in this group. Another staining reagent is *Coomassie blue*. DFO is used as a fluorogenic reagent for bloody prints, especially on porous surfaces.

Fingerprints from adhesive-coated surfaces Fingerprint development on adhesive tapes is a challenge particularly in trying to solve terrorist-related cases. In 1981, a successful development with *Coomassie brilliant blue 250* was reported. Since then, more sensitive staining methods have been studied; the current technique is based on staining the sebaceous components of the fingerprint deposit by a solution of *gentian violet*. Latent prints appear as dark purple impressions against a lightly stained background. The use of 'Sticky Side Powder' has also been proposed, and its use is becoming widespread.

Research and Development

Although the use of latent fingerprint detection to associate an individual with a crime scene started over a hundred years ago, systematic research in this area only began in the 1970s, with the optimization of existing techniques for fingerprint development and the search for new methods for particular surfaces. The lag between the general scientific knowledge and the scientific back-up to fingerprint practitioners is illustrated by the history of the use of ninhydrin. Shortly after the discovery of the ninhydrin reaction with amino acids in 1910, chemists noticed that many common materials, including sweat, form blue-colored products upon reaction with ninhydrin. Since the introduction of paper chromatography in the early 1940s, ninhydrin has been routinely used to locate amino acids on chromatograms. Despite its use over the years and the oft-repeated admonition against touching chromatograms exposed to ninhydrin, only in 1954 was ninhydrin recognized as a latent fingerprint reagent.

The application of physicochemical techniques to the visualization of latent fingerprints is one of the most dynamic and prolific areas of present research and development in forensic science. It can be divided into three categories: improvement and optimization of existing methods; the search for new reagents and techniques; and the search for solutions to problematic surfaces.

Two groups in particular have been active in optimizing fingerprint techniques (see Further Reading list for details). The main line of research at present is the search for nonpolar solvents to replace the banned freon 113, optimal working conditions for the chemical reagents (PD in particular) and the constant improvement of electro-optical devices.

Alternate light sources are becoming lighter, more powerful, more versatile and safer. Instead of the large and cumbersome 18 W argon-ion lasers, there are now more powerful, tunable illumination devices that emit light from ultraviolet to red, and can be operated at crime scenes. In addition, mobile argon-ion laser laboratories have been constructed, thus solving the portability problem. Specific modifications in other techniques, such as vacuum cyanoacrylate, introduction of field devices and electronic gating of optical observation systems are currently being tested. In an attempt to increase the detectability of latent fingerprints by amino acid reagents the use of proteolytic enzymes has been suggested that break down the proteins present in fingerprint deposits, to amino acids that react in a much stronger manner.

In 1973, systematic research of fingerprint powders and their adherence mechanism was started. Nearly twenty years later, the common aluminum powder was modified by increasing the content of stearic acid, thus lowering background noise and improving contrast. Powder dusting is the simplest and most widely used fingerprint technique; it still requires comprehensive research to have a better understanding of its

mechanism and to improve both its sensitivity and versatility.

Modern research into new techniques started, perhaps, in 1954 with the introduction of ninhydrin to fingerprint detection. Other milestones in this category are: the adaptation of the physical developer and vacuum metal deposition techniques; the introduction of lasers to fingerprint visualization and the secondary treatment with zinc salts; and the introduction of cyanoacrylate. Systematic research of chemical alternatives to ninhydrin was first reported in 1982 with the synthesis of several ninhydrin analogues. One of them, benzo[f]ninhydrin, showed some promise but was surpassed by other newly developed reagents. Other groups started to explore this area, and numerous such compounds have been prepared and studied. Of particular importance among them were 5-methoxyninhydrin and 5-methylthioninhydrin. The luminescence of their products with latent fingerprints after zinc chloride treatment was much stronger than that of ninhydrin-developed prints. The next step was the development of DFO, which has become the most important fluorogenic reagent for latent fingerprints. Current research on fluorogenic reagents focuses on substances from the 1,2-indanedione series (Fig. 5).

The search for solutions to problematic surfaces focuses on two surfaces in particular: human skin and spent cartridge cases. Fingerprint detection on human skin is a most difficult challenge since fingerprint deposits, as well as skin coating, comprise the same secretory compounds. Hence, in terms of 'signal to noise' ratio, all physicochemical methods give very poor results. In the few reported successes in death investigations, the latent prints were most likely contaminated with foreign materials such as oil or grease. Several groups, however, have reported successful trials under controlled conditions. Dusting with lead powder followed by X-ray imaging has been recommended. Among the other approaches more recently studied are initial transfer to glass, plastic film, or glossy Kromekote paper and further development by traditional methods; transfer to electrostatically charged plastic sheet and examination by oblique lighting; and an iodine–silver plate technique which starts with iodine fuming of the area under examination, transfer to a polished silver plate and exposure to strong light. A brown image of the lifted print is formed on the silver plate. Prints have been developed from cadavers by this method up to 14 hours after deposition.

Cyanoacrylate fuming, followed by treatment with the fluorescent dye, TEC, gave very promising results on postautopsy cadavers. The successful use of iodine-fuming followed by fixation with naphthoflavone has also been reported. However, attempts to apply the reported methods in actual cases gave significantly inferior results to those obtained under controlled conditions.

From critical evaluations of the various techniques it is quite clear that, for all practical purposes, there is no general solution at present to fingerprint development from human skin.

Fingerprint development from spent cartridge cases could be of great benefit in serious crime investigations. Most fingerprints, however, are destroyed during the shooting process. In 1995, it was shown that the main cause of deterioration is not high temperature as previously assumed, but the friction between the cartridge case and the chamber at the moment of ejection, which is caused by the widening of the cartridge due to high pressure during firing. Several groups have reported, however, that some portion of the fingerprint residue does remain on cartridge cases even after firing.

In 1976, latent fingerprints on spent cartridges were visualized by exposing them to nitric acid fumes and even by dusting with black powder. Positive results have also been achieved by cyanoacrylate fuming, especially on nickel cases, and by Gun-blue solution on brass, nickel and copper cases. It has been claimed that vacuum cyanoacrylate followed by fluorescent staining and treatment with selenous acid are suitable methods for fingerprint development on spent cartridges, and it has been shown that sebum-rich latent marks are much more resistant to the firing process. In some cases, they could be visualized by plating the cartridges with palladium or silver in solution or by exposing them to gold or aluminum vapors.

It is now widely accepted that the swelling stage and ejection of the firing process in semiautomatic weapons inhibits successful development of latent fingerprints on most types of ammunition. It is assumed that in those cases where positive results have been achieved, there was some tolerance between the cartridge and the chamber, so that even after the expansion, the friction was not too damaging.

See also: **Fingerprints (Dactyloscopy)**: Visualization; Sequential Treatment and Enhancement; Chemistry of Print Residue.

Further Reading

Almog J and Springer E (eds) (1995) *Proceedings of the International Symposium on Fingerprint Detection and Identification*. Israel: Israel Police.

Bridges BC (1963) *Practical Fingerprinting*. New York: Funk & Wagnalls.

FBI Academy (1987) *Proceedings of the International Forensic Symposium on Latent Prints*. Quantico: Virginia.

Forgeot R (1892) *Des Empreints Digitales etudiees au point de vue medico-judiciaire*. France: Document de Criminologie et de Medecine Legale.

Kent T (ed.) (1998) *Manual of Fingerprint Development Techniques*. London: Police Scientific Development Branch.

Kuno Y (1956) *Human Perspiration*. Springfield, IL: Charles C. Thomas.

Lee HC and Gaensslen RE (eds) (1991) *Advances in Fingerprint Technology*. New York: Elsevier.

Margot P and Lennard C (1994) *Fingerprint Detection Techniques*, 6th edn. Lausanne: University of Lausanne.

Menzel ER (1980) *Fingerprint Detection with Lasers*. New York: Marcel Dekker.

Olsen Sr RD (1978) *Scott's Fingerprint Mechanics*. Springfield, IL: Charles C. Thomas.

Police Scientific Development Branch (1974) *Proceedings of the Conference on The Science of Fingerprints*. London: Home Office.

Scientific Research Development Branch (1984) *Proceedings of the International Fingerprint Conference*. London: Home Office.

Shani A and Almog J (1982) *Methods for Fingerprint Development on Various Surfaces*. R & D Report DM/0604. Israel: Israel Police. (in Hebrew).

United States Department of Justice (1984) *The Science of Fingerprints, Classification and Uses*. Washington: FBI.

FIRE INVESTIGATION

Contents

Chemistry of Fire
Evidence Recovery at the Fire-scene
Fire-scene
Fire-scene Patterns
Laboratory
Physics/Thermodynamics
Types of Fire

Chemistry of Fire

P J Thatcher, Forensic Services, Northern Territory Police, Fire and Emergency Services, Darwin, Northern Territory, Australia

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0526

Introduction

The phenomenon of fire is a complex scientific event. It has had many simplified definitions applied to it in an attempt to provide a simple explanation for those who require a working knowledge of fire for practical purposes. From a forensic science viewpoint, this applies to those who are required to investigate and understand the ignition and propagation of fire, i.e. fire (cause) investigators. This knowledge must include a basic understanding of fire chemistry. This article outlines the basic chemical concepts that apply to combustion. It also defines numerous physical and chemical properties of gases and liquids that are necessary for their ignition and combustion and also explains their 'performance' on undergoing combustion.

A simple but scientifically accurate definition of fire (combustion) is as follows.

> Fire is an exothermic, self-sustaining chemical reaction involving a solid, liquid or gas-phase fuel and it is usually

associated with the oxidation of this fuel by atmospheric oxygen with the emission of energy in the form of heat and light.

It is evident then, that a fire requires fuel and oxygen and produces energy and oxidation products. However, unless the fuel undergoes autoignition whereby it does not require an external heat source, an external ignition source (piloted ignition) must be supplied before a self-sustaining exothermic chemical reaction will commence. This is the basis of the simplified 'fire triangle' (**Fig. 1**) that appears in all fire investigation texts and, when the need for a chemical chain reaction is included, leads to the more complicated 'fire tetrahedron' model (**Fig. 2**). If a side of the 'fire triangle' or a face of the 'fire tetrahedron' are not present then combustion can not occur. Simply, the physical requirements for fire are outlined below.

Fire

Conditions for fire

Air – Oxygen As mentioned in the definition of fire, the oxidizing agent is usually atmospheric oxygen. As will be explained later, the actual presence of oxygen is usually easily explained by the fact that it constitutes approximately 20% of the atmosphere. However, the amount of oxygen available is a critical parameter in the spread and intensity of a fire.

Under favorable conditions, oxygen can be supplied to a fire by the presence of other oxidizing agents, for example, potassium chlorate ($KClO_3$) and sodium nitrate ($NaNO_3$), which contain oxygen in their chemical structure. Also, under rare conditions, combustion can occur in an atmosphere of carbon dioxide or other gas in the absence of oxygen.

Fuel From a practical viewpoint, the fuel referred to in the definition can be regarded as any material that exists in a chemical state in which it can be oxidized by oxygen in the presence of a suitable ignition source.

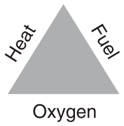


Figure 1 The fire triangle.

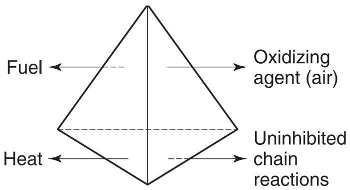


Figure 2 The fire tetrahedron.

The common fuels requiring consideration in fire investigation are organic compounds usually containing significant amounts of carbon (approx. 50% and higher) and hydrogen. They include naturally occurring compounds such as wood, cotton, etc., synthetic materials such as plastics, paints, rubbers, etc. and refined fuels and solvents which include petrol, lighting kerosene and methylated spirits.

Heat Energy in the form of heat is necessary to excite the molecules of both the fuel and the oxygen to an activated state required for chemical reaction. The minimum temperature necessary to initiate the 'self-sustaining chemical reaction' referred to in the definition of fire, is known as the ignition temperature of the fuel. It is the source of ignition that is the primary interest of fire investigators because this can be interpreted as the cause of the fire.

Fire requirements and processes

When a fuel and an oxidizing agent are heated to ignition temperature, either through autoignition or piloted ignition, chemical reactions occur. These exothermic (heat-producing) reactions provide additional energy to the situation and, as a consequence, further reactions proceed. When the production of excited species exceeds the decay of these species, the chemical reactions can be regarded as a chain or self-sustaining reaction and a fire or explosion results. The mechanism for this process is further discussed under 'Flame Chemistry' below.

The fire will continue until one of the reactants or the heat providing the energy source for continued ignition is expended. This forms the basis of methods for the containment and extinguishment of fire. Examples of fire-fighting principles include removing the fuel load, removing the oxygen source by covering a fire in foam or an inert gas and removing the heat supply by hosing with water on the principle that the latent heat of vaporization of water accounts for the heat.

Although combustible organic compounds exist in the solid, liquid or gaseous phase, most of these fuels will only burn when in the vapor phase, that is, a combustible gas-phase mixture must exist. For many volatile liquids this gas-phase mixture exists at ambient temperatures, but for others the liquid must be heated to a temperature at which the liquid will exert a vapor pressure which, when mixed with the required amount of oxygen, can support combustion. This is known as the flash point of the fuel. Simply, the flash point is the lowest temperature at which a liquid produces a flammable vapor. This is not to say, however, that the liquid will ignite if heated to this

temperature; it will still require an ignition source at the required ignition temperature. Furthermore, the flash point temperature will not sustain a fire. In practice, combustion requires a slightly higher temperature known as the flame point, to achieve this.

Although the flash point determines the minimum temperature at which a gas–air mixture meets the requirements for ignition, there is also a maximum temperature, above which, a liquid will exert a vapor pressure that is so high, that the required amount of air or oxygen will not be available for ignition. Therefore, there exists both an upper and lower flash point for liquid fuels. Most fire investigation texts report only the lower flash point because, in an actual fire situation, uncontrolled ventilation renders the upper flash point of academic interest only.

The preceding discussion refers to the need to evaporate a liquid, that is, to convert the liquid into the gas phase. A phase change from solid to liquid or liquid to gas requires an energy input referred to as latent heat. In the latter case, this is known as the latent heat of vaporization and is defined as the quantity of heat absorbed by a substance in passing between liquid and gaseous phases. A flammable liquid with a high latent heat of vaporization will require a high heat input to convert it into the combustible gas phase and, as a consequence, evaporation rates are limited and the heat is unavailable for other combustion processes such as pyrolysis. An example of a substance with a relatively high latent heat of vaporization is ethanol. Most are familiar with the cold sensation when methylated spirits comes in contact with the skin. This is due to the absorption of body heat in the vaporization process. The latent heat of vaporization is, therefore, another in a series of chemical and physical properties of liquids that determine their effectiveness as an agents to assist the ignition and spread of fire, i.e. to act as an accelerant.

Solid organic fuels, including natural products, such as wood, and synthetic products, such as plastics and rubbers, have a chemical structure that is based on high-molecular-weight molecules. In the case of wood, these materials include lignins and the naturally occurring polymeric carbohydrate cellulose, whereas in the case of plastics, synthetic polymers based on monomers, such as styrene and vinyl, are present. If these materials are to burn, they must first be thermally degraded by a process known as pyrolysis, to small molecules that exist in the gas phase.

Pyrolysis occurs when sufficient energy is applied to break chemical bonds, thus reducing large molecules to low-molecular-weight compounds. In many cases, these smaller compounds include monomers and dimers that were used to construct the polymeric compounds undergoing thermal decomposition. These pyrolysis products should not be confused with free radicals (see 'Flame Chemistry') that are also produced during thermal decomposition, but they will themselves become a source of free radicals.

The processes of evaporation and pyrolysis explain the general observation that flames occur above the surface of a solid undergoing flaming combustion rather than at the surface. There are important exceptions to this process and these occur when carbon (char) and some metals, e.g. aluminum and magnesium, burn. In these cases, there is direct oxidation at the surface of the solid and the process is normally very efficient, that is, the products of combustion can not be further oxidized. In the case of burning carbon, the process is referred to as glowing and, when complete combustion occurs, the combustion product is carbon dioxide. However, if the oxygen supply is restricted, incomplete combustion may occur with the production of carbon monoxide; a gas which is often responsible for deaths in fires. The subsequent combustion of this gas to carbon dioxide is characterized by a blue flame above the 'glowing' solid.

If combustion (i.e. oxidation) is 100% efficient, the following chemical equation applies for hydrocarbons (compounds containing carbon and hydrogen only) regardless of their origin:

$$\text{Hydrocarbon} + \text{Oxygen} + \text{heat} \rightarrow \text{Carbon dioxide } (CO_2) + \text{Water } (H_2O) + \text{Heat}$$

For toluene, a major component of petroleum, the equation is

$$C_7H_8 + 9O_2 \rightarrow 7\ CO_2 + 4H_2O + \text{Heat}$$

However, for uncontrolled combustion, where there is limited or no control over the supply of combustion reactants and removal of combustion products, the following equation more accurately defines the situation.

$$\text{Hydrocarbon} + \text{Oxygen} + \text{heat} \rightarrow \text{Carbon Dioxide} + \text{Water} + \text{Heat} + \text{products of incomplete combustion}$$

(carbon monoxide, soot, pyrolysis products, polynuclear aromatic hydrocarbons, etc.)

Again, for toluene, with a limited supply of air, the equation could be

$$C_7H_8 + 6O_2 \rightarrow 3CO_2 + 2CO + 4H_2O + 2C\ (\text{Carbon}) = \text{soot} + \text{Heat}$$

It is the products of incomplete combustion that form the smoke at the scene of a fire. Smoke is the cloud of soot particles, particulate matter, unburnt gases and combustion gases which, as a result of convection, rises above the fire. However, even with efficient combustion there may be gray 'smoke' seen in the gas plume due to condensing water vapor. The pro-

ducts of incomplete combustion may themselves be subsequently oxidized in a poorly ventilated fire. If such a fire is suddenly ventilated, these products may reach their explosive limits (see later) and ignite explosively in a process known as 'flashback' or 'smoke explosion'.

Flame Chemistry

Fundamentally, the investigation of fire origins and causes requires an understanding of fire-travel indicators such as depth of burning, lowest point of burning, etc. 'Glowing' fires, although extremely efficient oxidation processes, do not contribute significantly to fire spread, particularly early in the development and spread of fires and are usually associated with the middle and final stages. Most of the effect of fire can be attributed to the convection and radiation of heat from flames and, as such, an understanding of the basic chemistry of flames, which are the source of that heat, is where the science of fire investigation actually commences.

The various physical properties and parameters of gaseous fuels, the formation of combustible gases from liquids and solids and the requirements for ignition and combustion have been discussed. Also, the formation of oxidized and semioxidized products of combustion have been briefly reviewed. It remains to consider how this oxidation actually occurs in flames and why energy in the form of heat and light is produced.

Basically, the energy contained in any combustible material is a direct consequence of the chemical bonds that bind the material together. It is the breaking of these chemical bonds that releases this energy and oxidation is one chemical process or reaction that will achieve this. The amount of energy released in this way, i.e. by means of an exothermic chemical reaction, depends on the chemical composition of the reactants and the chemical composition of the products. (This is not to say that all chemical reactions are exothermic. On the contrary, many chemical reactions are endothermic and require the input of energy to allow the reaction to proceed and form products which have chemical bonds of higher chemical energy than the reaction products.) This release of energy occurs within the flame zone and, although flame structures and compositions are extremely complex, energy released can be explained simply as a consequence of the formation and properties of free radicals.

Free radicals are formed through the breaking of molecular bonds and are described as molecular fragments possessing one or more unpaired electrons. Because unpaired electrons are extremely susceptible to pairing with other unpaired electrons to form chemical bonds, these species are very reactive and short-lived. Any combination of free radicals is associated with the evolution of energy as the species combine to form a more stable entity.

During a fire, gaseous fuel (that could also be present due to an evaporated liquid or a pyrolyzed solid) forms a region rich in fuel vapor. Where this region comes in contact with an oxidant-rich region the gas phase combustion zone forms which is characterized by a thin flame sheet. Here the fuel is dissociated into extremely reactive free radicals that immediately react with atmospheric oxygen (and nitrogen) to form the final combustion products which, depending on fuel and oxidant conditions, are at various levels of oxidation. The energy released by these oxidation processes is radiated in all directions, including back to the fuel source and this insures that the chain reaction continues and the flaming area spreads. For small steady flames characterized by burning candles, cigarette lighters, etc., these flames are known as laminar diffusion flames. For larger unstable fires, they are known as turbulent diffusion flames.

There are some important consequences of free radical formation for fire suppression and fire investigation.

1. The removal of these free radicals provides an additional method of fire suppression and this has been utilized by fighting fires with chemical agents and chlorinated hydrocarbons.
2. The amount of energy released by fuels can be calculated. This can provide valuable explanations for the speed of fire spread and/or the intensity of damage. This value is known as the heat of combustion or heat output of the fuel. A simple definition for the heat of combustion is the maximum amount of heat that can be released by the complete combustion of a unit mass of combustible material.

This value can then be used to calculate a theoretical adiabatic (no heat gain or loss from the system) flame temperature that can provide information concerning the effect a flame can have on materials in the immediate area. There is, however, a further consideration before fire-spread can be predicted. This is the specific heat of fuels in the immediate area. The specific heat is the amount of thermal energy required to raise a unit mass of a substance 1°C.

Because fire is dependent on fuels attaining ignition temperature, the capacity of a fuel to absorb heat is yet another critical parameter in a consideration of fire spread. It effectively indicates the potential of a fuel to reach a dangerous condition or state during exposure to a fire and it is therefore an important consideration in fire suppression and fire-pattern in-

terpretation. The converse property, i.e. the resistance of a material to absorb heat, is known as the thermal inertia of the substance. Specific heat varies over a considerable range for common fuels but is less than $1\ \mathrm{cal\,g^{-1}\,{}^{\circ}C^{-1}}$.

Another important consequence of exothermic reactions is that, for an increase in ambient temperature, there will be an increase in the reaction rate. This gives rise to the concept of a Q_{10} value for chemical reactions. The Q_{10} value is defined as the increase in the rate of a chemical reaction that results from a temperature increase of 10°C. Generally, the Q_{10} value exceeds 2, which effectively means that, for every rise of 10°C the rate of the reaction doubles. Although this value assists in explaining why fires appear to grow at a rate that exceeds a simple time–development relationship, it would suggest that fires should rapidly become a conflagration. The fact that this does not happen is due to the limits on fuel availability and the increasing heat loss to the surroundings.

As previously stated, flames are primarily responsible for fire development and spread. Under uncontrolled conditions, flame combustion is a less efficient oxidation process than combustion which results from 'glowing'. This is easily confirmed by observations which reveal that a range of partially oxidized combustion products, such as soot, pyrolysis products and charcoal formation, are present with flaming combustion but are not seen during 'glowing'. In fact, the simple presence of a yellow flame often indicates that incomplete combustion is occurring because the flame color is due to the presence of incandescent carbon particles.

Explosions

The mechanisms of explosions are examined in detail in other sections in this text. However, because they represent a particular form of combustion they require some consideration here.

An explosion is effectively a rapid increase in pressure due to the formation of gases during combustion or detonation and/or the rapid heating of atmospheric gases. If this is to occur in the burning of hydrocarbons in the gas phase, certain conditions must be met.

Fuel–air (diffuse) explosions

In fires, the burning rate is controlled by the amount of oxygen available, that is the ventilation of the fire, and the fuel available for combustion. The supply of fuel is itself dependent on the rate of burning and therefore on the rate of pyrolysis. Therefore, to a certain extent, the production of flammable gases and their mixing with air is dependent on the fire itself. In an unconfined area and while there is adequate ventilation, products of combustion are removed from the system, air enters the system and the combustion process continues. However, in the case of explosions, the air and vaporized fuel are intimately mixed before ignition and, as a result, the ignition is extremely rapid. If this rapid combustion occurs in a confined space, then the resulting products of combustion and the existing atmospheric gases undergo a large and rapid expansion and the effects of an explosion are experienced. The pressure wave resulting from this rapid combustion is in the order of approximately a $1500\ \mathrm{kg\,m^{-2}}$.

Explosions are commonly associated with hydrocarbon vapors, but serious incidents also arise with other fuels in the form of vapors, mists, foams and dust suspensions. All these materials contribute to a type of explosion known as a diffuse explosion that is characterized by the presence of a combustible fuel–air mix in which the air is the oxidizing agent.

It was mentioned above that upper and lower flashpoints determine the temperatures at which vapors of liquid fuels can be ignited. It follows, therefore, that there must be particular concentrations of vapor and air which will support and sustain combustion. These concentrations are known as the explosive limits of the fuel. Simply stated, the explosive limits of a fuel are the upper and lower concentration limits of vapor in air which will allow explosive combustion. Because vapor pressures of combustible liquids decrease with increases in environmental or ambient pressure, flash points must increase with increasing pressure as a consequence.

There are, however, instances when explosions will occur although the fuel vapor–air mixtures do not fall within the upper and lower limits. If a fuel vapor–air mixture is too lean (insufficient fuel vapor concentration) and an ignition source is applied, this can result in localized heating which raises the localized fuel vapor pressure to a value within the explosive limits and an explosion will occur. Conversely, if a fuel vapor–air mixture is too rich (insufficient air concentration), the application of a heat source might cause enough turbulence to produce a localized fuel vapor–air mixture to fall within the explosive limits. Again an explosion can result.

For hydrocarbon fuels in a practical situation, the lower or lean limit often leads to an explosion only, with considerable disruption whereas the upper, or rich, limit is usually associated with an explosion and a significant fire. It is unlikely that the fuel vapor–air mixture would be consistent throughout an entire structure. It is more than likely that there are zones where the mixture is too lean to explode and other areas where it is too rich. It is equally likely that a zone exists between these extremes where the mixture

meets the requirements for an explosion. Should an ignition source be present in this area, a 'rolling' fire or explosion can result.

Technically, in the case of explosive combustion, there are three alternative processes. If the mixture of fuel and oxidant (generally, air) is homogeneous and the conditions for combustion are met, the chemical reactions occur throughout the entire reaction mass. This is known as a uniform reaction. Alternatively, if there is a clearly defined reaction zone (flame) separating zones containing predominantly fuel and zones containing predominantly oxidant, then the reaction zone (flame) will move through the reaction mass. This is known as a propagating reaction. In this case, the propagation velocity at which the reaction zone or flame moves, further defines this type of combustion. When the reaction zone velocity is less than the velocity of sound (i.e. subsonic) the reaction is known as a deflagration and when it exceeds the velocity of sound (supersonic) it is termed a detonation.

Dense-phase (concentrated) explosions

Dense-phase or concentrated explosions differ significantly in that they are not dependent on oxygen in the air but result from an oxidation made possible because the explosive compounds contain both the fuel and the oxidant. With the input of a small energy source, the chemical structure of these compounds is rearranged with the simultaneous production of much more energy. This energy can be measured and is known as the heat of decomposition. Common explosives in this class include dynamite, trinitrotoluene (TNT), pentaerythritol tetranitrate (PETN) and nitroglycerine. More details concerning the structure of explosives and the mechanisms of explosions are provided in other articles.

In most actual fire scene situations, many of the principles and processes mentioned always apply. However, the relative significance of many of these will vary with the actual situation and relatively insignificant changes in one or more of the fire conditions may have a profound effect on other processes and effects.

See also: **Explosives**: Mechanism of Explosion. **Fire Investigation**: Physics/Thermodynamics.

Further Reading

Barnard J and Bradley J (1985) *Flame and Combustion*, 2nd edn. pp. 113–129. New York: Chapman and Hall.

Bersch W, Holzer G and Sellers CS (1993) *Chemical Analysis for the Arson Investigator and Attorney*, pp. 20–23. Heidelberg: Hüthig Buch Verlag.

Cooke R and Ide R (1985) *Principles of Fire Investigation*. Leicester: Institution of Fire Engineers.

DeHaan J (1991) *Kirk's Fire Investigation*, 3rd edn, pp. 6–66. Englewood Cliffs: Prentice Hall.

Drysdale D (1985) *An Introduction to Fire Dynamics*, pp. 1, 33, 78–252. New York: Wiley-Interscience.

Luke RH and McArthur AG (1978) *Bushfires in Australia*, pp. 23–46. Canberra: Australian Government Publishing Service.

National Fire Protection Association (1991) *Fire Protection Handbook*, 17th edn. Quincy, MA: National Fire Protection Association.

Redsicker D and O'Connor J (1997) *Practical Fire and Arson Investigation*, 2nd edn. pp. 55–90. New York: CRC Press.

Weiss C (ed.) (1986) *Hazardous Chemicals Data Book*, 2nd edn, pp. 1–15. Park Ridge, NJ: Noya Data Corporation.

Yearance R (1987) *Electrical Fire Analysis*, pp. 141–145. Springfield, IL: Charles C Thomas.

Evidence Recovery at the Fire-scene

P J Thatcher, Forensic Services, Northern Territory Police Fire and Emergency Services, Darwin, Northern Territory, Australia
J D Kelleher, Fire and Explosion Investigation Section, Victoria Forensic Science Centre, Melbourne, Victoria, Australia

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0527

Introduction

Arguably the most critical point in a fire investigation is the interface between the fire scene investigator and the forensic science laboratory. This is because, in most instances, the scene investigator does not conduct the laboratory testing or present the laboratory results. This is particularly so in cases involving criminal proceedings.

As a consequence, in these cases the presiding magistrate or members of a jury must determine the significance of the test results in the context of evidence given by the scene examiner. Predictably, other evidence, such as motive, will also become available during the hearing but the scientific evidence concerning the actual starting of the fire should not require the support of other types of evidence. Obviously, other matters, such as who was responsible for the starting of the fire, will require other evidence.

The Purposes of Collecting Samples

The collection of samples and items at fire scenes can serve several purposes. Undoubtedly, most items are taken for subsequent examination by the fire scene investigator or laboratory scientists with a view to obtaining information to support ideas and opinions established during the scene investigation.

Although this is a justifiable procedure, it should be remembered that any results obtained from the testing do not necessarily prove the validity of the scene diagnosis. For example, positive results from tests designed to investigate a diagnosed ignition cause do not prove that this was the actual cause. These results merely confirm that such a theoretical cause is a practical concept. Furthermore, any negative results do not prove that the diagnosed cause could not have occurred. In these cases, testing would have to be repeated to ascertain the likelihood that a fire would commence in this way; that is, an investigation would determine the statistical improbability of this particular ignition source. So long as there is a theoretical possibility that a fire can start in a particular way, negative testing will only provide evidence as to how unlikely this is, not that it is impossible.

A classical example of this philosophical viewpoint in the interpretation of test results is fires in bedding that are caused by smoldering cigarettes. Simple tests using lighted cigarettes and cotton sheeting show that it is difficult (or a statistical improbability) to start fires in this way. However, it is well established that there are many deaths caused through smoking in bed. Therefore, the testing proves that, for every death that occurs, many people survive because cigarettes usually burn out in this situation.

It is obvious from the preceding discussion, that the quality and quantity of items and samples have the potential to impact profoundly on laboratory results and, therefore, on the fire-cause diagnosis. This article addresses the scientific considerations for sampling, including the necessary control and reference samples. Furthermore, because samples have such scientific significance, they can have evidential value in any subsequent criminal or civil hearings. Therefore, legal considerations, such as item security and item continuity, must be addressed in order to satisfy jurisdictional and legislative requirements.

Sampling from fire scenes also serves other purposes. Important examples include research and education into product performance under actual fire exposure conditions. The results from these observations can lead to academic research and investigations into product performance, safety and improvement.

The recording of the performance of materials subjected to fire under different conditions assists the fire investigator to 'read' the fire's progress and therefore to diagnose a fire origin or starting point. However, this product performance can also provide the necessary evidence required for a product re-call and also lead to recommendations for changes and improvements in national standards for materials used in building, furnishing, clothing etc.

Sample Classification

Before considering the scientific aspects of items and samples, it is important to classify samples according to their intended functions. This will allow identification of sampling requirements and of the limitations on information that these samples can provide. The following definitions apply and can, in fact, apply to items collected from scenes of incidents other than fire.

Test sample

A test sample is taken to provide specific information that will assist in reaching conclusions or establishing facts concerning the incident. Examples include clothing with biological stains, vehicle panels with paint smears and floor coverings containing flammable liquids. A feature of test samples is that they usually consist of a substrate containing a foreign material.

Control sample

A control sample is taken for comparative purposes and provides information that is often necessary to determine the significance of the results produced from test samples. This is achieved by sampling and analyzing the material that comprises the substrate of the test sample, and eliminating these results from the test sample results. Examples of control samples include carpet, floorboards, soil and other fire debris. The extraction of natural products such as lanolin from woollen carpet underlay, and terpenes from timber when collected as control samples, allows the scientist to account for these materials when interpreting the analytical results from test samples. Because these samples are basically used for elimination purposes, an important feature of control samples is that they can fulfill their purpose without the necessity of a complete scientific identification of their composition or of substances that occur naturally within them. Control samples are not always required but should always be considered.

Reference samples

A reference sample allows an absolute scientific or technical identification to be made of a test or control

sample, or of an extract from a test or control sample. Because a reference sample is used for comparison purposes, it might also serve as a control sample under certain circumstances. For any sample to serve as a reference sample it must satisfy one important criterion, that is, the material must have scientific and/or technical evidence of its identity and composition. The most common reference samples used by the laboratory in fire investigations are fuels and solvents provided by oil refineries. However, there might also be the need to examine reference samples of carpet when reporting and explaining the significance of burn patterns or the identity of pyrolysis products. The manufacturer would normally supply such samples but, providing comprehensive technical information is made available, a carpet supplier could provide the necessary samples.

Samples from Fire Scenes

As previously discussed, samples and items are collected for various purposes. However, by far the most common reason for collecting samples from fire scenes is to provide information concerning the ignition of the fire. Furthermore, the majority of these samples consist of fire debris suspected of containing flammable liquid residues.

Selection of sampling sites

The sampling sites for fire debris samples are usually selected after an interpretation of the 'fire travel indicators'. These allow the fire investigator to interpret the fire travel direction and intensity and, therefore, determine the fire origin or starting point. Whether the ignition of the fire was assisted by the presence of flammable liquids is decided after a consideration of various observations and other empirical evidence (see below).

Over the past 20 years, some fire investigators have, to a certain extent, avoided the orthodox scene inspection procedure and have resorted to the use of hydrocarbon vapor detectors known as 'sniffers' to 'screen' the scene for the presence of fuel and solvent vapors.

'Sniffers' have undergone development from the initial simple instruments based on a color change in crystals exposed to hydrocarbons, to sophisticated portable gas chromatographs. Fire investigators must be familiar with the specificity and sensitivity specifications of an instrument before using it in the field and herein lies the dilemma.

Many thermal breakdown products, known as pyrolysis products, are present in fire debris as a result of the thermal decomposition of natural and synthetic products. Many of these products are very similar in their chemical structure to aliphatic and aromatic compounds occurring in petroleum fuels and solvents. Therefore, a positive reading from a 'sniffer' should not be regarded as confirmation of petroleum compounds but only as an indicator of their possible presence.

Also, petroleum products often undergo severe evaporation when exposed to an intense fire. This evaporation will continue after the fire has been extinguished and before sampling occurs. In fact, it is not uncommon to recover residues of fuels and solvents that only consist of the highest boiling point components of the original mixture. These chemical compounds can exert such low vapor pressures that the 'sniffer' cannot detect them.

As a consequence of these limitations of 'sniffers', the fire investigator must make a decision based on experience as to whether a sample should be taken from a promising site, regardless of a negative response from the 'sniffer', and whether a sample should not be taken from an unlikely site regardless of a positive response from the 'sniffer'.

The question then arises whether the 'sniffer' is serving any purpose at all and, if not, why use it at all?

The situation is somewhat different when the 'sniffer' is a portable gas chromatograph. If the use of this instrument produces a chromatogram that is definitive evidence of a flammable liquid, then it is providing valuable investigative information that has the potential to become evidence in any subsequent court hearing.

There is no substitute for a competent and logical scene investigation when diagnosing fire origins. The presence of flammable liquids at a fire scene has some significance but the presence of a flammable liquid at the diagnosed fire origin is of greater significance.

Fire debris samples

Samples of fire debris usually consist of floor coverings, floorboards and soil but burnt clothing, curtains, soft toys, bedding, etc. are also collected and submitted for examination. If maximum information is to be gained from these items, it is essential to have at least a basic knowledge of the chemical structure of these materials and their burning characteristics. For the more common sampling media, the following information applies.

Floor coverings When these materials burn with the assistance of flammable liquids, there are usually characteristic 'pooling patterns' produced that are characterized by a sharp demarcation between burned and unburned material. However, it should be remembered that similar patterns can be produced

in other ways, for example, falling drapes, burning wallpaper and protection provided by furniture and fittings. In most cases, an experienced fire investigator will be able to distinguish the actual cause of any burning pattern and will sample accordingly.

If it is decided that sampling is warranted, the sample must include both burned and unburned carpet (that is, the sample should be taken at the edge of the burn) and any underlay must be included. Normally an area of at least 1 m^2 should be taken and this is best achieved using a large heavy-duty knife.

Floor boards So long as holes in flooring are not too extensive, the fire investigator should be able to distinguish the cause of these holes. Burning debris usually causes holes burnt from the upper surface, whereas holes burnt from the lower surface can often be associated with a flammable liquid. In the latter case, if the burning is not too extensive, the characteristic 'tongue and groove' burning patterns are often present. This pattern arises when the liquid seeps through the cracks in adjoining floorboards and burns up through this seam. This can result in a series of holes in the flooring that follow the joins of the floorboards. The pattern can occur without the involvement of flammable liquids, but its presence always demands a close inspection.

If it is decided to take a sample, the sample must include the unburned material on each floorboard in addition to the charred areas. These samples should always be accompanied by control samples because wooden flooring usually contains natural products that have the potential to complicate any laboratory analysis. Sampling is best achieved using an axe or a saw. If a power saw is used then serious contamination considerations arise if the saw is fueled by a petroleum product.

If the flooring has been burned to such an extent that the holes are almost joined, it is extremely unlikely that any flammable liquid residues will remain and sampling is probably worthless.

Soil In many instances where the involvement of a flammable liquid is suspected, but floor coverings and flooring are too severely charred to warrant sampling, consideration should be given to sampling the soil underneath the floor. In these cases, the top few centimeters of soil should be removed for testing using a clean spade or trowel. Again, a control sample should be taken, although it will probably not be used because extracts from soil are often relatively free of contamination.

For all the above cases, the taking of control samples is a consideration. They should be collected from areas that are confidently diagnosed as not being involved in the starting of the fire. In many instances they will not require subsequent scientific examination, but their existence can sometimes be of great assistance to the laboratory.

There may also be the opportunity at the scene to collect important reference samples that will assist the laboratory to identify extracts from the test samples. Importantly, these samples might also provide important incriminating evidence in investigations and any subsequent criminal proceedings.

Sample size

Sample size is obviously determined by individual circumstances. If a section of carpet or other fire debris is saturated in flammable liquid, then the sample size required by the laboratory will obviously be significantly less than for a sample that has no odor and a diffuse burning pattern or is extensively charred. Sample sizes for control and reference samples will not normally be as large as those for test samples where the amount of sample extract is the major consideration.

As a rule, samples cannot be too large but they can be too small. If a sample is inadequate, it may not be possible to return to the scene and re-sample. Often, the laboratory analysis is not conducted for some time after sample submission and the fire scene could be cleared during this interval. Also, even if the scene does remain, it is unlikely that there will have been any scene security since the original inspection and the integrity of any further samples is therefore compromised.

Packaging

Over the years, many different types of packaging and containers have been used to collect and store samples that contain fire debris and flammable liquid residues. Basically, these containers fall into the following classifications.

Metal containers Drums and cans have been used for many years. They are light, strong, cheap, effectively retain flammable liquid vapors and can be used during the analysis of their contents. However, they are bulky and, when the varnish coating on the interior is scratched, they are quickly susceptible to rusting. Unfortunately, fire debris samples usually contain nails, broken glass, rocks, etc; all of which easily damage the varnish layer. Also, there are usually copious amounts of water in these samples from the fire fighting operations and so corrosion commences immediately. Therefore, any samples collected and stored using metal cans should be analyzed as soon as possible by the laboratory.

Bottles Glass bottles have also been used for many years with reasonable success. Depending on the type of lid seal used, they are generally effective in retaining flammable liquid vapors for long periods. They can also be washed and reused, although this is not advisable and unnecessary given the cheap cost. However, glass bottles are heavy and fragile and additional care must be used in their handling.

Plastic bottles have been used to overcome the deficiencies of glass bottles. They can be effective so long as the plastic is impervious to hydrocarbon vapors.

A major disadvantage of all bottles is the restrictions placed on the sample size by the size of the bottleneck and lid.

Plastic bags Paper bags are obviously totally ineffective for collecting samples. However, the fact that samples are still occasionally submitted using these bags suggests that this is worth noting here.

'Polythene' (polyethylene) bags have been used with limited success. These bags are cheap, light, are available in a variety of sizes, are durable and can be easily sealed in the field by tying at the neck. Unfortunately, they are not impervious to hydrocarbon vapors and can be affected by some solvents and chemicals. The use of these bags should be regarded as temporary packaging at the best.

The shortcomings of these bags can be overcome by the use of polyvinylidine bags (available as 'oven' or 'freezer' bags). These bags have been shown to be impervious to hydrocarbon vapors and will contain flammable liquid vapors for periods in excess of a year under reasonable storage conditions. However, a major disadvantage with these bags is that, if the bag is pierced, a tear usually results very quickly.

Because all these packaging types have certain advantages and disadvantages, no particular method is recommended. Rather, a combination of methods should be considered depending on the circumstances and availability of the containers. For example, samples could be placed into 'polythene' bags to take advantage of their resilience and then placed into cans or polyvinylidine bags to retain any flammable liquid vapors.

Sample hygiene

The recovery of samples from areas of interest often requires tools and equipment that are not normally present in a laboratory. These tools include pinch bars, large knives, shovels, axes and saws. Many fire investigators also include a small chain saw with their equipment.

This equipment must be thoroughly cleaned after each sample has been taken, to prevent contamination of any subsequent samples. This can be done using detergent or some other cleaning agent that does not contain substances with a petroleum or ethanol base.

For similar reasons, if a chain saw is included in the equipment, extreme care must be taken to avoid contamination of the samples by the fuel for the chainsaw.

Number of samples

As discussed previously, control and reference samples should be taken liberally although, in many cases, they will not require examination. The number of test samples collected will depend on the perceived potential of the samples and the individual requirements of the jurisdiction. For example, many jurisdictions are served by fire investigators with little practical or theoretical training but who have extensive experience. When these investigators are unsure as to the fire origin and fire cause, they often take many samples with the hope that the laboratory will detect a flammable liquid residue in one or more of these samples and so identify the fire cause and origin. This approach is actually the reverse of accepted fire investigation procedures that require the origin to be diagnosed before sampling commences to investigate the cause. Also, as a consequence, large amounts of laboratory time are used to examine samples taken with no logical basis and the results obtained are open to various interpretations.

If it is obvious from burn patterns and odors that flammable liquids have been present in a number of rooms in a building, it should not be necessary to sample from all of these sites. Rather, a high quality sample should be taken that will undoubtedly confirm the presence and identity of any flammable liquids, and accurate notes taken of the odors and burn patterns in the other rooms.

Other items and samples

With the vast number of diagnosed ignition sources, there are a corresponding variety of items that require collection and examination. In many cases, test samples will not be taken because the fire has destroyed all the physical evidence in the vicinity of the fire origin. However, in many of these cases, the collection of control samples will be vital if any information is to be salvaged. An example of just such a situation is a fire that is ignited in bedding, either deliberately or by a smoldering cigarette. Often in these cases, no bedding remains to be examined and the actual ignition source is completely destroyed. However, vital information will still become available if identical bedding is available for laboratory testing. In these situations, reference samples of the bedding might

also be required, especially if the burning characteristics of the bedding become the major issue.

When the ignition cause is diagnosed as a piece of machinery or electrical component, a subsequent laboratory examination can often provide supporting information. In these instances, it can be advantageous if similar components that have also been exposed to the fire can be examined in order that the significance of the test sample results can be determined. Effectively, the additional components act as control samples.

Electrical fuses, fused electrical wiring and other electrical components can provide much information during an examination at the scene and, in fact, may critically influence the scene examination and final fire cause diagnosis. If it is considered that further information could become available through a more detailed examination in the laboratory, control samples should also be collected. In the case of fused wires and 'blown' fuses, samples of unaffected and fire-affected components should be submitted as control samples.

There are occasions when no test samples are available and no control or reference samples will assist the investigator. A classical case is a bush fire (wild fire) caused by a lightning strike. In these cases, even if the particular tree struck by lightning survived the ensuing fire, nothing will be achieved by examining the tree in the laboratory that could not be achieved by examining the tree *in situ*. Photographs recording the condition of the tree will be required, as a formality, but the actual fire cause diagnosis will depend more on prevailing weather conditions and burning patterns in the immediate area.

Sample Labeling

All samples must carry appropriate identification through adequate labeling. Agency and legislative requirements can influence what is entered on labels but, as a general rule, the following information should be included.

The label must include details of who took the sample (including a signature) and the time and date the sample was collected. Ideally, the label will also include a basic identification of the sample and the site from which it was taken, for example: 'Sample of carpet from near the settee in the north east corner of the loungeroom', Sampled by N. Collins, 1400 hrs, 31/11/98, signature.'

This labeling fulfills two functions: (1) it will allow the results to be interpreted by the fire investigator, magistrate, coroner, jury, etc.; and (2) it will allow a positive identification to be made of the sample at a later date. This latter point is significant when it is realized that any subsequent court proceedings may be years after the incident.

Basic labeling can be applied at the scene and consolidated when the items are submitted to the laboratory at a later date. Obviously, when a sample is labeled, the site of the sampling must be entered onto the scene diagram in the scene case notes.

Sample continuity

Adequate labeling can also assist in the establishment and maintenance of the records necessary to prove sample continuity. If a court subsequently requires any item, or result obtained from an item, collected at a scene, continuity must be strictly maintained and proven. In effect, sample continuity or integrity means that the presence of the sample must be accountable at all times. This must be documented and, if necessary, established through written or oral evidence. Evidence bags are available with sealing and labeling facilities that allow all persons who have access to the bag to record the fact on the bag. However, statements will still be required if it is intended that these items become exhibits.

Summary

The recovery of evidence from fire scenes should be based on observations, interpretation and an appreciation of the laboratory's requirements and capabilities. As laboratory examinations are time-consuming and expensive, it is inefficient, wasteful and potentially damaging to a case to use the laboratory for the primary diagnosis of a fire cause. The laboratory can, however, provide valuable supporting information for a diagnosed ignition source if the necessary test, control and reference samples are collected, preserved, labeled and submitted.

See also: **Crime-scene Investigation and Examination**: Recording; Collection and Chain of Evidence; Recovery of Human Remains; Packaging; Preservation; Contamination. **Evidence**: Classification. **Fire Investigation**: Types of Fire; Physics/Thermodynamics; Chemistry of Fire; Fire-scene.

Further Reading

ASTM Standard E1188-95. *Standard Practice for Collection and Preservation of Information and Physical Items by a Technical Investigator*.

ASTM Standard E1459-95. *Standard Guide for Physical Evidence Labelling and Related Documentation.*

National Fire Protection Association (1995) *NFPA 921: Guide for Fire and Explosion Investigation*. Quincy, MA: National Fire Protection Association.
National Fire Protection Association (1997) *Fire Protection Guide to Hazardous Materials*, 12th edn. Quincy, MA: National Fire Protection Association.
Sax NI (1975) *Dangerous Properties of Industrial Materials*. New York: Van Nostrand.
Weiss G (ed) (1986) *Hazardous Chemicals Data Book*. New Jersey: Noyes Data Corporation.

Fire-scene

R H Ide, Forensic Science Investigations, Sutton Coldfield, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0528

Introduction

The appearance and the evidential value of all items and materials at the fire scene will have been influenced by three separate sets of conditions:

1. the nature of the building and its contents;
2. the behavior of the fire;
3. the fire-fighting and subsequent activities.

It is normally the task of the investigator to establish details of the behavior of the fire, where and how it had started and why it had spread in a particular manner. The fire investigation is therefore an inquiry into the unknown set of conditions sandwiched between two, which can under normal circumstances easily be established.

When the investigator has established the particular objectives for the current inquiry, the decision must then be made as to whether the structure is safe enough for a full investigation to proceed. Evidence may not be suitable for recovery immediately, but all steps should be taken to preserve both physical and factual evidence. During the investigation, all information obtained should be recorded in the most effective manner possible, and electronic and manual methods are likely to be used at most fire scenes.

Building and Contents

Although the structure of a building plays an important part in determining how and whether a fire develops, the materials used in most buildings are relatively difficult to ignite and, as a result, it is almost always the contents which play by far the greatest part in the initial development of the fire. It is normally possible during an investigation to establish what items are alleged to have been present within the building and, indeed, photographs, video recordings and detailed witness evidence may be readily available. In addition to this the residual materials found after the fire may give clear evidence as to the true nature of the furnishings or contents of the building. Undamaged rooms or sections of rooms may provide relatively unburned samples, which may be analyzed, identified, or even used for small-scale tests.

It may therefore be possible to form strong opinions as to the probable fire behavior, if the alleged contents of the building had been in the positions described by the last legitimate occupant prior to the fire.

Certain types of soft furniture can be easily ignited by a small heat source, such as a lighted match or even a cigarette. Most modern furniture is manufactured to more rigorous standards and is therefore less easily ignited. If the fire had developed in vulnerable soft furnishings, and particularly if there had been the opportunity for the fire to spread up a vertical surface, then very rapid fire development could have occurred from an early stage in the fire. Fire spread from this point would depend on the proximity of other flammable materials and, even more, on the available ventilation.

Witnesses or other evidence may be available to establish whether doors or windows were open or closed, particularly if the doors had been fitted with intruder alarms. Detailed information as to the types of doors fitted and the ways in which they had been secured may be available from the owners or occupants of the building. This witness information should be compared with the evidence available at the scene, taking into account the possibility that a subsequent intruder may have opened or closed doors. Windows may also provide evidence of this type, although it may be more difficult to establish whether or not a window had been broken prior to the fire. Other factors of significance are whether or not the windows had been double glazed, and how effectively the building had been insulated.

The diligence with which the information regarding the original state of the building is pursued will be affected by the objectives of the investigation. Less information relating to the state of the building, security of windows, flammability of building materials and so on, will be required if the objective is simply to establish where and how the fire started. If, on the other hand, the main force of the inquiry is into

reasons for particularly severe damage or allegations that the contents of the building had not been as disclosed by the owner, then more thorough inquiries will be necessary.

It is self-evident that information regarding building contents should not be accepted at face value, because the informants may have a vested interest in giving misleading information to the investigator. In addition, the last legitimate occupant of the building may not have truly been the last person inside the building, and it is possible that a fire setter had been present and had re-arranged the contents of the building to assist fire spread or had deliberately introduced fire accelerants.

Behavior of the Fire

Although the basic principles of fire development are well understood small changes in the environment may have large effects on the final result. For example, a fire developing in a situation with limited ventilation is likely to develop slowly, or even self-extinguish, whereas a fire in a building with broken or open windows may develop extremely rapidly.

If the principles of fire development are well understood, then the investigator may be able to use a system of reverse logic to establish how and where the fire started. For example, fires show a strong tendency to spread upwards, rather than downwards and the lowest point of burning might be thought to be the seat of the fire. The situation is somewhat complicated by the possibility of falling burning materials during the development of the fire and the effect of heat radiated downwards, once the flames have reached the ceiling. During the course of the fire, there are normally many witnesses present and these may be able to provide assistance interpreting the evidence found.

Firefighting and Subsequent Activities

It is unfortunate that the interests of a firefighter and a fire investigator do not fully coincide. The role of a professional firefighter is to save life and property. For this reason a rapid, determined and resolute approach is necessary to carry out these functions before the occupants of the building are incapacitated or killed by toxic gases, oxygen depletion or heat. Because of the rapidity with which a fire can develop it is necessary for a firefighter to act quickly, by forcing entry to the building in the most expedient manner. It will be necessary to remove smoke and toxic gases by ventilating the building, and to extinguish the fire as rapidly as possible.

Forcible entry

It is common for firefighters to force doors, and rapid methods are used preferentially, irrespective of the damage caused. Techniques include kicking the bottom of the door, using a sledge hammer or axe near to the lock or using some form of levering device, designed specifically for the purpose. Professional firefighters will normally be able to provide information identifying which doors had been forced by them, and which doors had been found forced upon their arrival. It may not, however, be possible to distinguish between doors forced by illegal intruders and those forced by well-meaning passers-by in an attempt to rescue the occupants of the building.

Searching

If any of the occupants of the building are unaccounted for, firefighters wearing breathing apparatus will search the building in an attempt to locate and rescue any trapped survivors. During the search, which is likely to be carried out in dense smoke, articles of furniture and other contents of the building are likely to be disturbed. Articles disturbed at this stage in a fire should be recognizable from their physical appearance, even if the firefighters do not have an accurate recollection of the exact items moved. It is normal practice in most brigades to recover bodies from buildings, even when they are apparently dead, because of the possibility that they might be resuscitated after their removal to a safer, fresh air environment.

Venting

It is common practice for windows and doors to be opened and broken during firefighting in order to clear smoke from the building as rapidly as possible. However forced ventilation is increasingly being used. This technique, known as positive pressure ventilation (PPV), can clear the smoke from a building very rapidly and, used in conjunction with water, may very much increase the speed of the firefighting activities. However, it is likely to produce abnormal burning patterns, particularly at a high level, which may mislead investigators, unless they are aware that the technique has been used.

Extinguishment

By far the most common extinguishing medium used in almost all fire situations is water. This may be employed in the form of jets or sprays, depending on the size and intensity of the fire being fought. Fire appliances incorporate pumps capable of producing

jets of water with sufficient force to dislodge plaster, break windows, overturn furniture and drive debris beneath carpets. When used effectively, firefighting water generates large volumes of steam which, in addition to having a cooling and smothering effect, also produce a pressure rise which can drive the fire front away from the firefighter. For this reason, it is desirable, although not always practical, for firefighters to be stationed between the fire and unburned parts of the building.

Movement of burned items

It is of great importance for firefighters to insure that all burning materials have been extinguished before they leave the scene of the fire. For this reason, debris is soaked with water and portable burning items capable of smoldering are removed from the building. Ceilings may be brought down, wall panels forced away from the walls and floor boards pulled up in order to ensure that no hidden residual seats of fire remain.

These precautions can result in the loss of significant evidence for the fire investigator. Some brigades instruct their firefighters that in cases of suspected arson, they should use sprays rather than jets and reduce the disturbance of debris to a minimum, particularly at the supposed seats of fire.

Fire brigade information

Because of the methods used by fire brigades, the firefighting may have more effect on the final state of the scene than the fire itself. It is, therefore, of great importance that the investigator should, in relevant cases, obtain information relating to the techniques used by the firefighters and take into account for example the fact that certain parts of the building may have been inaccessible and may have burned for more prolonged periods than others. These regions will show high levels of destruction, but may not be the point of origin of the fire.

Fire Investigation Objectives

Before the investigator commences work, the objectives for the investigation should be established. Although it is normal for one of the requirements to be to establish the cause of the fire, there may be other objectives depending on the needs of the organization commissioning the investigation. For example, in the investigation of a fire which behaved in an unusual way or caused considerable loss of life, it may be necessary to identify factors which contributed to the loss, such as dangerous materials present or materials behaving in an unusual and unexpected manner. Common reasons for investigations include: detecting arson, identifying dangerous appliances or processes, establishing liability, identifying reasons for damage, identifying dangerous materials and considering reasons for deaths.

Safety

It is important that the investigator should consider the safety of the building before entering or instructing others to enter. Although possible physical dangers may be apparent upon external examination of the building, dangers due to chemicals present within the fire scene may only become evident as a result of information provided by the owners of the building.

It is important that the investigator should learn to make judgments as to what level of damage to brickwork can be regarded as safe and what structural damage should be regarded as posing a threat to safety. Loose slates and tiles, separated bricks at a high level and broken glass from skylights may, although loose, appear to be relatively safe if undisturbed. However, under windy conditions or if many people are working in the building, loose items may become dislodged and fall from a significant height presenting a serious danger to those working below. Damage to walls may be minor, but cracks to brickwork or stonework may develop during the period after the fire. It is important, therefore, to establish whether cracks, bulges and other damage to walls and structural members appear to be growing more severe during the period of time following the fire.

Injuries may also occur to investigators as a result of damage to floors and stairways. It is possible for holes to be concealed by debris, for apparently strong boards to have been badly damaged from beneath and for synthetic board floors to become soft and weak when wet. It is desirable to inspect any flooring of questionable strength from underneath if possible.

Chemical dangers

In any industrial premises there are likely to be chemicals or manufacturing materials, which may present particular safety problems. Solvents, raw materials, chemical intermediates and acids may frequently be encountered in industrial premises. Although the works chemist or process engineer may be aware of the chemicals involved directly in their manufacturing process, they may not be aware of chemicals held within equipment or within the structure of the building.

For example, PCBs (polychlorinated biphenyls) may be present in electrical equipment and asbestos may be found within the heated portions of appliances and beneath the roofs of some buildings as a sprayed layer. Carbon fiber is used as a reinforcement in certain materials, beryllium oxide is sometimes used as an electrically insulating ceramic, and chromium copper arsenate has been used as a wood preservative. In a destructive fire, chemicals of this type may be released as fine respirable dusts and all of these can constitute serious health hazards.

In addition, many relatively harmless materials can produce dangerous chemicals when heated. In the case of polyvinyl chloride (PVC) most of the harmful chemicals are dissipated with the smoke of the fire, but other compounds, such as carcinogenic pyrolysis products, and hydrofluoric acid from certain fluorelastamer rubbers, may reside within the partially burned and redeposited tarry materials at the scene.

Biological hazards too may also exist at a fire scene. It is by no means certain that the microorganisms within a dead animal or human will be killed by the heat of the fire and in some cases it is possible that the elevated temperatures may have caused such organisms to multiply. It would normally be prudent to assume that any dead body or carcass removed from a fire scene is carrying pathogenic organisms. For this reason gloves and other protective clothing should be worn when handling or moving a body or carcass.

Action

Take advice from as many valid sources as possible, including fire officers, building inspectors, the management or owners of the premises and other investigators. If the building is particularly dangerous, it may be necessary to carry out the investigation after it has been demolished or made safe.

Awareness

When at a fire scene it is necessary to be alert at all times and to be watchful for your own and for others' safety. Where possible, it is desirable to know the locations of other workers in the building and it is to your advantage that they should know where you are. Any changes in circumstances, such as an increase in wind speed, apparent movements of the building, sudden falls of debris or a discharge of dust from a prominent crack, may be warning signs that the building is becoming increasingly unsafe and should be evacuated.

Clothing

Protective clothing should be worn at all serious fire scenes. Such clothing would include a helmet, protective overalls, strong boots and one or more pairs of gloves. It will sometimes be necessary to wear a respirator designed to prevent the inhalation of fine dust.

Safety plan

A safety plan should be made and adhered to. It may become evident that certain parts of the building can be investigated safely, but that other parts would be too dangerous for any investigator to enter. It is necessary to be aware of the known hazards in the building, particularly any structural problems, and of the escape routes to safe points within the fire scene and, above all, to be aware of the possible chemical and biological dangers, together with the precautions and the procedures necessary to counteract their effects.

Preservation of Evidence

Although considerable destruction of evidence will have taken place during the fire, and subsequent firefighting, much is likely to remain. However, this can steadily disappear as a result of disturbance, evaporation, partial dislodgement of the building materials and the removal of items. The most serious evidence destruction of all is caused by ill-advised, premature, unscientific investigations. It is possible that evidence may be present but that it has not yet been recognized by those attending the scene. Partly for this reason, the first stage of the investigator's examination should include a widespread examination of the fire scene and surrounding area, taking comprehensive photographs and recording information.

During the preliminary investigation, the investigator will be seeking the point or points of origin of the fire and attempting to assess how it had spread to cause the resultant damage. In addition, he will be seeking possible points of entry to establish whether any unauthorized person had succeeded in entering the building. Evidence of forcible entry, if rescue and firefighting attempts can be ruled out, must indicate that an intruder had been present. However, it is possible to start fires in buildings without physically entering the premises and the introduction of an ignition source should also be considered.

Other possible information may be available from addressable intruder or fire alarms, video surveillance equipment and from witnesses. Part of the task of preservation of evidence includes recording information from witnesses or at least recording their addresses so that information can be obtained subsequently.

Although it is important to preserve the evidence directly relating to the fire scene, it is sometimes the

case that material in the vicinity may also provide information useful to the investigation. For example, fire setters have been known to discard items in the undergrowth near to the attacked building. The items may include ignition sources or empty containers that had previously held accelerant.

Discarded items may also relate to the contents of the building. For example, drawers, bags or boxes, stolen with their contents from the building, may be discarded empty by the intruder when he feels he has reached a place of safety. Items discarded by an intruder may provide fingerprint evidence, and the possibility of finding significant footwear impressions in the vicinity should be borne in mind.

Recording of Information

For many reasons it is common for fire-damaged buildings to be restored or demolished relatively quickly after a fire. Any examination carried out by an investigator is unlikely to be repeatable. The investigator has, therefore, a responsibility to make a comprehensive record of the findings. Photographs taken from as many angles as possible of all regions of the fire scene which appear relevant can be of great value.

However, apparently irrelevant parts of the building may later become significant as the inquiry develops it is important to take photographs as comprehensively as possible. One of the great problems of fire investigation is that with the blackening of the walls and contents of the building, the disruption of the electrical supply and possibly the sooting-up of windows, illumination is normally very limited. A camera loaded with a fast film, equipped with electronic flash, may reveal items and evidence not clearly visible when the investigator was at the scene. It can be useful to examine the photographs carefully, as soon as possible, in case significant evidence has been revealed which may require further analysis at the scene.

Camcorders

The use of camcorders at scenes is becoming increasingly widespread. In common with still photography, a thorough coverage of the scene is desirable. A video recording has the advantage that different parts of the scene may be seen in association during a gentle panning shot. The investigator may choose to deliver a commentary whilst recording, although there is the possibility that undesirable comments made by others present at the scene may also be recorded. With most camcorders it is possible to insert a blank jack plug into the external microphone socket to silence the recording.

Plans

Much information showing the relative positions of items can be recorded in the form of plans. It is common practise to draw a rough plan, including measurements, and to construct a more accurate plan under more favorable conditions. If this procedure is followed, it is necessary to preserve the original plans, no matter how dirty and poorly drawn they may be, because they constitute the original notes, which may be required in a court of law. When plans are drawn the investigator must make a conscious decision as to whether the plan should represent the situation when the scene was examined, the situation believed to have prevailed at the time of the fire, or the situation when the scene was left by the last legitimate occupant. All three of these situations will be different. Items may have been removed from their habitual place by an intruder and subsequently disturbed again during firefighting. It is desirable that the plan should record the positions of doors and windows and, where possible, the situation of switches and taps.

Notes

The investigator is likely to make written notes in addition to the photographs and plans. These notes can often conveniently be recorded on predesigned forms intended to act as *aides-mémoire* to ensure that all significant aspects of the investigation have been addressed. An alternative method of making notes is to dictate them, using a small handheld tape recorder. Although this procedure is convenient, particularly in conditions of heavy rain or snow, it is necessary to preserve the tape as original notes for production in court.

See also: **Fire Investigation**: Types of Fire; Evidence Recovery at the Fire-scene; Fire-scene Patterns.

Further Reading

Cook RA and Ide RH (1995) *Principles of Fire Investigation*. Leicester: Institution of Fire Engineers.

Cardoulis J (1990) *The Art and Science of Fire Investigation*. St. John's: Breakwater.

DeHaan JD (1997) *Kirk's Fire Investigation*, 2nd edn. New York: John Wiley.

Munday JW (1994) *Safety at Scenes of Fire and Related Incidents*. London: Fire Protection Association.

Noon R (1995) *The Engineering Analysis of Fires and Explosions*. Boca Raton: CRC Press.

White P (1998) *Crime Scene to Court. The Essentials of Forensic Science*. London: Royal Society of Chemistry.

Fire-scene Patterns

R H Ide, Forensic Science Investigations, Sutton Coldfield, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0529

Introduction

Examination of the physical and chemical evidence left at a fire scene may, provided that there has not been too much disturbance, provide evidence which may enable the cause of fire to be recognized, the point of ignition to be identified and the types of fuels first ignited to be established. Directional effects may be clearly evident and the causes for particular smoke staining patterns may often be deduced. The melting or discoloration of materials within the building will provide some evidence of the temperatures attained at particular locations within the fire. To some extent, fire scene patterns can assist in the recognition of the types of fire involved. Differences between slow burning fires and rapidly developing fires can normally be recognized and these may often be further categorized.

However, there is a considerable body of unsubstantiated mythology associated with the examination of fire scene patterns and some investigators make assertions about the behavior of charred wood, cracking glass and soot, which simple experimentation would have clearly shown to have been incorrect. For example, rules of thumb based on the depth of charring of wood should be treated with the deepest of suspicion. Some investigators still believe that all wood chars at a constant rate of 1/40 in min^{-1} (0.06 cm min^{-1}). Simple tests demonstrate the intuitively obvious fact that the rate of charring depends on the amount of radiant heat acting on the piece of wood.

Preliminary Examination

It is a common experience for fire investigators, when first viewing a severely damaged building, to see the situation as hopeless and unlikely to yield any significant evidence. However, very considerable bodies of evidence may be recovered after even the most severe of fires.

The investigator will immediately upon arrival at the scene begin to form opinions as to the proportion of the building damaged and will be starting to try to assess which parts of the fire damage occurred during the early stages of the fire and which occurred later. Photographs taken during the fire-fighting by the brigade and by on-lookers may help in this assessment, although they are unlikely to be available during the initial stages of the investigation. Clearly, it is normal for fires to spread from low parts of the structure upwards through the building, and if there is any significant wind, it too is likely to have played its part in influencing the direction of spread. There is likely to be more severe damage on the downwind side of the seat of the fire, even to the extent that sometimes greater heat damage may be found on the downwind side of an item rather than on the side nearest the seat of the fire.

The fire-fighting, too, is likely to drive the flames away from the initial points of attack, whilst suppressing burning near to the regions of application of water. Positive pressure ventilation, a technique increasingly used by fire brigades, is also likely to result in unusual heat and smoke patterns. This technique involves the use of powerful blowers, which force fresh air into the building, displacing smoke. Some increased burning at high levels in regions remote from the seat of the fire may occur.

High-level investigations

At some stage in the investigation, it will be necessary to decide how much importance to attribute to evidence at a high level in the fire-damaged building. Although it is unlikely that the main point of ignition will be at a high level, in cases of deliberate ignition, there may be separate ignition points at different levels in the building. In addition, it is possible for fires to spread downwards, normally by the fall of burning material from a point of ignition at a higher level.

In general, however, the most easily investigated seat of fire is likely to be at a low-level within the burnt part of the building and it is possible that any evidence that had originally been present at a high level, will have been destroyed by the flames. Point of entry evidence is also unlikely to be evident at high levels in the building. It is, however, desirable for the investigator to have a pair of binoculars available, so that evidence visible in inaccessible parts of the building may be examined safely. If the building is unsafe to enter, some examination can still be carried out from a hydraulic platform using binoculars and a camera with a telephoto lens. It is also possible that the highest parts of the building were the places where victims had been trapped by the fire, and it may be necessary to recover bodies or to examine the places where they had been found.

Exterior walls

It is normally desirable to make an exterior examination of the building and of the smoke-staining which is often apparent on the walls above windows

and doorways. This staining may provide evidence of forcible entry and of windows and doors which were the first to fail during the fire. The smoke-staining to the walls may indicate the direction of the wind prevailing at the time of the fire, although it should be borne in mind that the wind direction could have changed during the period that the building burned. Normal experience of bonfires and barbecues demonstrates that the direction of smoke travel is likely to change during relatively short periods of time.

Point of entry

In many deliberately ignited fires, the premises will have been forcibly entered by an intruder, and such evidence may be clearly visible to fire-fighters and subsequent investigators. If the intruder ignited the fire from within the building, it is probable that he would have started the fire at some point remote from his point of entry, in order to escape safely as the fire developed. For this reason, it is not uncommon to find significant evidence at points of forcible entry even after the fire and the fire-fighting, provided that reasonable care has been taken by the fire brigade to preserve this valuable evidence.

Work has been carried out on techniques of restoring fingerprints left on glass and other smooth materials, subsequently affected by smoke, heat and water. In addition there may be footprints on window sills, tables and other articles of furniture near to points of entry. Other evidential materials, such as glass, blood and fibers, may also have survived the fire and fire-fighting at the point of entry. In many cases of arson by intruders, far greater implicative evidence is likely to be found at the point of entry than at the point of ignition. It should also be borne in mind that many of the more experienced criminals tend to ensure a point of exit for themselves and this region is even more likely to be free from fire spread damage.

Location of Point of Ignition

Fire scene patterns are among the methods which may be used to locate the point or points of ignition. In general, these techniques depend on the assumption that where the fire has been ignited, it is likely to burn for the longest time, producing the greatest amount of damage. Although in simple fires this may be a valid assumption, in complex fires involving difficult fire-fighting, unequal distribution of fuels or variable ventilation throughout the building, then fire scene patterns can produce misleading evidence for fire seat location and more reliable and impartial techniques should be used. Pattern-based techniques which have been commonly used include the following.

Seeking the region of maximum damage within the building Although this technique may be fairly reliable in the case of small fires, in large fires the maximum damage can be very much influenced by ventilation, the available fuel, and the fire-fighting techniques used.

Depth of charring to wood The rate of charring of wood will depend on the type of wood involved and the amount of radiant heat flux to which it has been subjected. The greater the heat flux the more rapid the rate of charring. This effect can be used to establish where the greatest heat has been involved, by comparing charred wood in different parts of the building. However, in all fires there is likely to be a considerable thermal gradient, with higher temperatures being attained at high levels within rooms. For this reason, the most meaningful comparison will normally be made at a low level where differences in temperature are more significant. The depth of charring does not provide any absolute measurement of the time for which the timber in question had been subjected to heat.

Spalling of plaster When plaster is subjected to heat it may spall from the underlying brickwork of the building (**Fig. 1**), providing some evidence which can be used in fire seat location. However, the sudden cooling effect due to fire-fighting water jets may also spall plaster from brickwork, although in this case there is unlikely to be any smoke deposition on the surface of the bricks.

Distortion to glass and plastic Many plastics soften at temperatures of 130–400°C. The manner of their distortion can give indications as to the direction and magnitude of the heat acting on the original item.

Figure 1 Plaster spalled from a wall as a result of cooling by fire fighting jets. The direction of the spray can be seen from displaced soot on the ceiling.

Glass tends to soften at 750–900°C. Above 900°C it is sufficiently mobile to fall to a lower, cooler position.

Distortion of metalwork Structural steelwork in buildings distorts in severe fires and the degree of distortion may indicate the regions where the greatest heat had been released.

Melted metal Commonly encountered metals melt at temperatures between 326°C for lead and over 1000°C for metals such as copper and steel. Evidence of significant quantities of melted metal may indicate that very high temperatures had been achieved in certain parts of the building.

Damage to concrete floors Concrete spalls at elevated temperatures, and high temperatures at a low level within the building may be indicated by spalling to concrete floors. Unfortunately, because of the variation in quality of concrete likely to be encountered in many buildings, confusing damage may occur. Contrary to general belief, spalling to a concrete floor does not necessarily indicate that a flammable liquid had been burnt on the surface and in fact during the burning of a flammable liquid the wetted surface tends to remain cool until all of the liquid has evaporated.

Residual heat The region where the greatest burning had occurred is likely to have been heated more than any other part of the building. It has been argued that, after the fire has been extinguished, the regions where the greatest residual heat remains may represent the seats of the fire. One complicating factor is that during fire-fighting water applied to suppress the fire will not necessarily cool the whole building uniformly and so the residual hot areas may simply represent areas where less fire-fighting water was deployed.

Burning of soot Although soot deposits in different ways (to be discussed later) one indication which has been used for fire seat location purposes, is evidence that soot has been burnt from surfaces after first being deposited. This indicates an elevated temperature and an oxidizing atmosphere and may be of significance in the enquiry.

Low burning Since fires spread predominantly upwards, the lowest region of burning might reasonably be supposed to be the seat of the fire. However, burning materials can drop down and start other fires and, in addition, a well-developed fire may possibly burn through a floor, causing damage at a lower level.

Holes in the floor When there is a single hole in the floor of a building, this may well be the seat of the fire. However, there are other possible explanations, for example it might be a region of particular fuel loading, a region of particularly good ventilation, a region where the floor had been weaker than normal or a region last to be extinguished by the fire brigade. If liquid fire accelerants have been used, they may leak through joints between floorboards causing characteristic burning patterns. However, burning beneath and between floorboards can be produced by normal fire spread. This damage cannot be regarded as conclusive evidence of the use of fire accelerants unless supported by analysis information.

Funnel patterns A seat of fire may produce a characteristic funnel pattern either in smoke or heat damage in the room or building involved (**Fig. 2**). However, funnel patterns are not necessarily categorical evidence of seats of fire because regions of particular fuel loading or regions where burning materials have dropped can also produce similar patterns.

Characteristic structural collapse Buildings may collapse in a manner characteristic of a fire having been started in a particular region. Interpretation of such evidence should be carried out with considerable caution because widespread travel of hot air and smoke may produce conflicting effects, as also may unequal distribution of forces within the building.

Thermal direction indicators If the furniture and contents of the building have been largely left in position, then charring and heat damage to particular faces of the furniture may provide an indication of the

Figure 2 A characteristic funnel pattern indicating an armchair as the point of ignition. The fire had, in fact, been ignited approximately 500 mm from the chair, but had spread to it at an early stage in the fire.

direction of the seat of the fire. It is important that the original positions of items should be known categorically for this type of evidence to be used.

Ceiling damage Particularly intense damage to one region of the ceiling may indicate a seat of fire beneath it. However, it must be recognized that a fire at a low level in a room will cause widespread and general heating to a ceiling and the damage at a high level is less likely to be diagnostic than intense damage at a low level.

When the fire-spread pattern information and the other more objective techniques of ignition point location have been employed, it is likely that the investigator will be able to define the probable seat of fire to within certain limits. It should be possible to draw a line defining the outer limits for the possible position of the point of ignition, the confidence perimeter or radius of error. Excavation within the area defined by the confidence perimeter must include the seat of the fire. It is clear, therefore, that the whole of this area must be excavated during the investigation if any physical evidence of the cause of the fire is to be found. In general terms the more rapidly the fire has developed, the larger the confidence perimeter, because a rapidly developing fire tends to produce relatively uniform damage.

If a fire started slowly, for example by smoldering, then very localized damage may be found. It is, therefore, possible by examination and comparison of the damage to form some opinion as to how rapidly the fire has developed. In a fire which has started by smoldering and burnt slowly until it was extinguished, characteristic localized deep charring may be found. The smoke-staining, characteristic of a slow-burning fire, may show evidence of nonbuoyant pyrolysis products causing staining on the horizontal upper surfaces of objects in the vicinity.

A rapidly developing fire, on the other hand, can cause relatively uniform heat damage, with smoke which is likely to be buoyant and probably more carbonaceous than that from a smoldering fire (**Fig. 3**). A complication may arise because a fire which started in the smoldering form may develop into a free-burning flaming fire at some stage, possibly when the ventilation is increased, causing confusing evidence. In the most extreme case, the transition from smoldering to a free-burning fire may take the form of a 'back draught'. This phenomenon occurs when flammable gaseous pyrolysis products accumulate within the building under conditions of limited ventilation. The opening of a door or window may allow the flammable gases to ignite explosively.

The reverse of this process may occur if a large fast-burning fire is imperfectly extinguished and an area of smoldering remains unnoticed by the firefighters. Under these circumstances, a smoldering fire can occur causing intense burning in one region. Such regions of burning are likely to be in inaccessible places, in the floor boards, at the base of the wall, at the sides of cupboards and under or behind furniture. Such regions of damage, which are known as 'bulls' eyes', are relatively easy to recognize.

Figure 3 Typical smoke-logging in a passageway. Hot, buoyant smoke produces staining at a high level.

Information from Smoke Staining

The appearance of deposited smoke at fire scenes can provide valuable evidence, giving information relating to the position of items, their temperature, the behavior of the fire and the direction of smoke travel. Although the composition of smoke can vary considerably, the buoyant black smoke produced in a free-burning fire tends to deposit preferentially on cold items rather than hot, and rough surfaces rather than smooth and it deposits better when flowing in a turbulent manner, rather than laminar.

The tendency for smoke to deposit on cool surfaces is often clearly apparent, resulting in effects of selective deposition related to the underlying materials beneath the surface in question. Electrical wiring,

structural steelwork and even heads of nails may act as preferential areas for smoke deposition under certain circumstances (**Fig. 4**). Surfaces coated with soot may suffer a rise in temperature during the development of the fire and particularly if this rise in temperature is associated with increased availability of oxygen; then areas of previously deposited soot may be burnt away producing a clean area in the middle of a blackened region.

Deposits caused by smoke from smoldering fires tend to be different in appearance, consisting of sticky, tarry pyrolysates, which tend to deposit very much more evenly, although there is a pronounced tendency for the smoke to fall, lacking as it does the buoyancy of the carbonaceous smoke from a free-burning fire.

Recognition of Types of Fire

Fire scene patterns may provide evidence to indicate the type of fire involved. Fires accelerated by the use of flammable liquids may show characteristic pool burns, particularly if they are extinguished at a relatively early stage. A pool burn shows a hard-edged demarcation line, between the irregularly shaped liquid pool and the remaining unburnt floor covering. Because of the lowering of the surface tension of most commonly encountered flammable liquids when they are heated, liquid accelerant tends to be driven under the unburnt region of floor covering. After a destructive liquid fire, samples taken from this transition zone are more likely to contain recognizable accelerant residues than samples from other regions.

Although a liquid-accelerated fire may well be evidence of arson, it is possible for flammable liquids legitimately present on the premises to cause similar effects. In addition, certain synthetic carpeting materials may burn in a manner which mimics the effect of a pool burn, even if no liquid accelerant had been present. For this reason, it is normally essential that samples should be analyzed to demonstrate the presence of a liquid fire accelerant.

Figure 4 Smoke tends to deposit preferentially on cooler surfaces and may reveal positions of nails and joints normally hidden by the plaster.

Fires caused by smoking materials

Although smoking materials are commonly blamed for many fires, it is difficult to demonstrate that a fire has been caused in this way. Carelessly discarded smoking materials include not only discarded lighted cigarettes, but also discarded lighted matches and tapers. Although a discarded lighted cigarette can normally only initiate smoldering, a carelessly discarded lighted match or taper can start a fire which shows all the characteristics of a free-burning flaming fire from its inception. If a lighted cigarette is suspected of having caused the fire, very localized damage is likely to be found at the point of ignition. The position of this localized damage must have been within material which was capable of smoldering and which could have been ignited by a small ignition source. Bedding, soft furnishings and piles of rubbish are the materials most commonly ignited by cigarettes. Normally, smoldering can only occur in cellulosic materials capable of forming a rigid char. Upholstery foams are not easily ignited in this way unless a cellulosic fabric is also involved. In addition, evidence that the occupants of the building had been careless with smoking materials is likely to be found, with cigarettes, spent matches or spent tapers scattered carelessly in the unburned parts of the building.

Recognition of spontaneous combustion

In most cases of spontaneous combustion, the development of heat is likely to have been slow, and the maximum temperatures are likely to be attained within the bulk of the material which has self-heated. Although there will evidently have been disruption of the material during fire-fighting, there may still be evidence that the interior of the pile of material has reached higher temperatures than the exterior. The heating will occur in a susceptible substance, such as hay, cotton, fishmeal, sawdust and coal. There is also likely to have been a background history of heating prior to the fire.

More reactive substances such as oil-soaked rags, car body filler, quick-drying paint and certain metals are likely to self-heat more rapidly and can do so in smaller quantities. With the more reactive substances, the evidence of internal heating may not be so clearly defined as in the less reactive substance, where a larger quantity must necessarily be present for ignition to occur.

Recognition of electrical fires

The fire scene pattern evidence for electrical fires is confined to the recognition that the seat of the fire is in, or associated with, an electrical appliance or system. For an electrical appliance to cause a fire it must necessarily either emit sparks or flames, or reach a temperature such that its exterior surface is hot enough to ignite materials in the vicinity. Under both of these circumstances it is to be suspected that the interior of the appliance will have been significantly hotter than its surroundings. This is normally apparent on examination, but the increasing use of beryllium in electronic components makes the examination of electronic devices which have achieved very high temperatures extremely dangerous, because of the possible risk of inhaling beryllium oxide dust. Such examinations should only be carried out under laboratory conditions, where stringent precautions can be taken against this serious danger.

Evidence of electrical arcing in a fire scene does not, in itself, provide evidence that the fire was electrical in origin, although it may be taken as evidence that the conductors in question were live at some time during or after the fire. However, an electrical appliance suspected of having caused the fire must firstly have been live and must secondly have been in the region where the fire started. For this reason, damage to the insulation of wires within or supplying the appliance is likely to result in electrical arcing, which may have caused the operation of a fuse or miniature circuit breaker. The significance of arcing within the appliance can therefore be assessed accordingly.

Recognition of fires caused by lightning

When a fire occurs in a building during an electrical storm, the suspicion may arise that it had been caused by a lightning strike. In most countries records are kept of individual lightning strikes and it is possible to establish from a meteorological department whether such a strike had occurred in a particular area, at a particular time. In addition, examination of the fire scene patterns may demonstrate damage to roofing tiles and electrical arcing effects to vertical wiring. The effects may be particularly characteristic in wires leading between telephone inputs and receivers and vertical wires to television aerials. In addition, plug sockets may have been disrupted explosively by the lightening strike and there may be particular damage to television sets. Such evidence may well survive after relatively destructive fires.

Recognition of arson

A high percentage of fires in many countries is attributed to deliberate ignition. Although it is possible to cause a serious fire by the application of a flame to a single artifact within a building, it is more common in cases of deliberate ignition for there to be several points of ignition and for liquid fire accelerants, such as petrol or paraffin to be used. The presence of extraneous liquid fire accelerants provides strong evidence that the fire has been ignited deliberately and their presence may be suspected if liquid pool burns are found or traces of unburned liquid are smelt (**Fig. 5**).

Multiple seats of fire must be shown to be independent of one another if found and in a large fire it may be difficult to establish whether apparently separated seats of fire were simply the result of differing firefighting effort or different amounts of fuel present in the regions of the building. If it can be conclusively demonstrated that there had been several separated seats of fire then, provided that accidental reasons for multiple seats can be excluded, then the fire can be concluded to have been deliberately ignited. Accidental reasons for multiple seats include secondary electrical faults, disturbance of burning material, interrupted critical processes, and abnormal fire spread.

Figure 5 A typical petrol vapor explosion, which caused almost complete demolition of the building, followed by a widespread and destructive fire.

See also: **Fire Investigation**: Types of Fire; Physics/Thermodynamics; Chemistry of Fire; Evidence Recovery at the Fire-scene; Fire-scene; Laboratory. **Wood Analysis**.

Further Reading

Cook RA and Ide RH (1995) *Principles of Fire Investigation*. Leicester: Institution of Fire Engineers.

Cardoulis J (1990) *The Art and Science of Fire Investigation*. St. John's: Breakwater.

DeHaan JD (1997) *Kirk's Fire Investigation*, 2nd edn. Chichester: John Wiley.

Munday JW (1994) *Safety at Scenes of Fire and Related Incidents*. London: Fire Protection Association.
Noon R (1995) *The Engineering Analysis of Fires and Explosions*. Boca Raton: CRC Press.
White PC (ed.) (1998) Crime scene to court. *The Essentials of Forensic Science*. London: Royal Society of Chemistry.

Laboratory

P J Thatcher, Forensic Services, Northern Territory Police Fire and Emergency Services, Darwin, Northern Territory, Australia
J Kelleher, Fire and Explosion Investigation Section, Victoria Forensic Science Centre, Melbourne, Victoria, Australia

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0530

Introduction

The effectiveness of any laboratory examination conducted on items taken from a crime (or other) scene is fundamentally linked to the quality of the scene examination. Included in this examination should be the recording of the scene, the interpretation of events through scene reconstruction, sample selection and sample packaging. The premise applies particularly to the laboratory examination of items taken from fire scenes.

In addition to diagnosing a possible fire origin and cause, a competent fire-scene examination can provide evidence of point of entry and other offenses. As a consequence, a thorough scene examination could require the attendance of a team of experts with a range of different skills, who have the potential to generate a range of samples requiring different scientific and technical examinations. In order that the maximum scientific value is obtained from these samples, the forensic science laboratory must provide scientific skills that complement the scene skills of the investigative team. Therefore, although this article will specifically refer to laboratory skills and techniques applied to fire-cause investigation, it is necessary to consider briefly the range of fire-scene examination skills provided by an investigative team.

Scene Expertise

- **Fire-cause investigator** A fire-cause investigator will attempt to determine the fire origin and cause. Supplementary information concerning ignition, fire spread, combustion properties, etc. will then be provided by the laboratory examination of samples taken from the scene.
- **Fingerprints examiner** A fingerprint expert can locate and recover evidence at fire scenes pertaining to the identity of victims and offenders. The expert often submits liftings of latent fingerprints and other items to the laboratory for further examination.
- **Crime scene examiner** A trained crime scene examiner will locate evidence that assists in the reconstruction of the incident and collect evidence accordingly, for example, castings of jemmy marks from the point of entry, shoe impressions, tire impressions, burnt documents, biological samples, etc.
- **Crime scene recorder** Although all members of the investigative team will meticulously record relevant aspects of the scene, occasionally specialists are required to take photographs, record videos and draft accurate plans.
- **Other experts** Some fires require the attendance of specialists from other disciplines. Examples of this include electricians, engineers and, in the case of fires in clandestine laboratories, drug analysis experts.

Laboratory Expertise

If it is expected that a single laboratory will provide full scientific support to the investigative team, a wide range of forensic science skills will be required in that laboratory. These will include the following.

- **Forensic biology** The examination of items containing biological evidence (blood, saliva, semen, etc.) can assist in the identification of an offender and/or victim through DNA profiling.
- **Fingerprints** Specialized techniques for the visualization of latent fingerprints can be conducted in the laboratory. Also, any fingerprints located can be classified and compared with fingerprint files for the purpose of identifying an offender or victim.
- **Criminalistics** The comparison of casts, photographs of toolmarks, etc. with items taken from suspects and offenders are conducted in the laboratory to assist with the identification of offenders.
- **Document examination** Document examiners can use specialized methods to restore, conserve and decipher fire-damaged documents. These skills can assist investigators examining the circumstances surrounding the fire.

Notwithstanding these important investigative requirements, by far the largest scientific effort in the investigation of fire scenes is the examination of items

collected for their potential to provide evidence concerning the fire cause. Although the majority of these examinations concern the extraction and identification of flammable liquids and flammable liquid residues from debris, clothing, soil and other samples, there are other less common analytical procedures that require consideration.

The remainder of this article will focus on all these issues.

Flammable Liquids

The identification of areas of a fire scene possibly affected by the presence of flammable or combustible liquids (see below), and of areas where flammable or combustible liquid residues may remain, are clearly relevant to a fire scene examination. Identification of these liquids can be of importance in determining fire spread and can be of critical importance in the interpretation of the fire cause.

It should be noted that flammable liquids, such as petrol, have flash points at or below 61°C. Combustible liquids, such as automotive diesel fuel, have flash points above 61°C. The term 'flammable liquid' generally includes combustible liquids unless otherwise specified.

Flammable Liquids of Petroleum Origin

In a commonly encountered fire-scene situation, most flammable liquids present are derived from petroleum products, typically petrol, kerosene, mineral turpentine or automotive diesel fuel. These liquids are not water-soluble, and they are absorbed by, or adsorbed onto, soil, furniture, floor coverings, fittings and building materials. They all evaporate at temperatures readily achieved in fires and some even exert significant vapor pressures at room temperature. A flammable liquid layer can be consumed if the substrate itself is consumed. However, because of modern fire-fighting techniques, in many cases substantial unburnt absorbent or adsorbent material remains, containing identifiable flammable liquid residues.

There are three steps in the identification process. The flammable liquid must be extracted, it may require clean-up, and finally, identification through analysis.

Extraction

Direct sampling This technique can be used for samples of flammable liquid, including relatively dilute samples. It can be used for both liquid and gaseous substances. With liquid samples, dilution with an appropriate solvent may be required, because most analytical methods are designed for low flammable liquid concentrations. Gaseous samples can also be collected directly. With sample containers suspected of containing large amounts of flammable liquid vapor, the headspace vapor can be collected using a syringe. The syringe will necessarily be of greater volume than that used for liquids, possibly 2–5 ml compared to the 2–5 μl used for liquid samples.

The direct methods are simple and relatively rapid, although they have limited application. They require minimal preparation time, and allow almost immediate analysis. The obvious disadvantage is their relatively low sensitivity, and also, in the case of vapor samples, direct gaseous injection tends to favor collection of lower boiling point components.

Headspace adsorption This is the recovery technique most widely used at present. In all adsorption techniques, the flammable liquid vapor is adsorbed onto a suitable medium, often activated charcoal or a proprietary adsorbent such as 'Tenax'. The adsorption process usually requires the sample to be heated and may be passive or dynamic. In the former, an adsorption medium is placed in the sample container and allowed to stand for some time, and in the latter, the headspace vapor is extracted through the adsorption medium by a pump.

Headspace adsorption is an extremely sensitive method, simpler and more rapid than steam or vacuum distillation and does not require expensive glassware or extensive equipment cleaning operations. However, the extreme sensitivity increases the potential for contamination and interpretation to become issues, and headspace techniques cannot easily be related to the total amount of flammable liquid in the sample. Headspace adsorption techniques generally favor recovery of lower boiling point hydrocarbon components.

The adsorbent is often enclosed in a glass tube that can be prepared as required or can be purchased prepacked. There are other forms of support in regular use, such as adsorbent mounted on a wire or paper strip. These are essentially minor variations, and the same considerations apply to these as to methods using glass tubes.

Solid phase microextraction (SPME) SPME is a recent development that is similar to headspace adsorption but without the necessity for the solvent extraction step. A fused silica fiber with a suitable adsorbent coating is inserted into the sample container to collect flammable liquid vapor from the sample headspace. The collected material is loaded onto a gas chromatograph column by a rapid thermal desorption process.

The limiting factor with solid phase microextraction is the efficiency of the adsorption process, as is the case with conventional headspace adsorption. This is a relatively new method which, in view of its simplicity and speed, may become more common. The sensitivity is comparable to that of headspace adsorption methods.

Distillation The oldest extraction technique is distillation. Steam distillation, where steam is passed through the debris to vaporize and remove the flammable liquid, is most common, although vacuum distillation has also been used. It is a relatively insensitive method, but it has the advantage that it can provide a guide to the amount of flammable liquid originally present. It requires relatively large samples but is less sensitive to contamination than some other methods. Distillation techniques tend to favor recovery of higher boiling point petroleum components.

A major concern with steam distillation is the recovery of water-miscible flammable liquids. Ethanol, in the form of methylated spirits, is a water-miscible solvent commonly used as a solvent and cleaning agent. The boiling point (78°C) of ethanol is low compared to most flammable liquids, and losses by evaporation can be substantial. More importantly, ethanol is miscible with the water distilled in the process, and, as a consequence, as the distillation progresses, the concentration of ethanol will decrease.

Therefore, in the steam distillation operation, where a water-miscible substance is known or suspected to be present, the first few milliliters of water distilled should be analyzed. At this stage of the process, the flammable liquid concentration will be at a maximum, and the distillate will not be significantly diluted. The difficulty of identifying these low-boiling-point, water-miscible substances is a significant argument in favor of alternative methods of sample treatment.

Solvent extraction This is a basic chemical technique that can be applied to the analysis of samples of fire debris. It is conveniently applied to small samples, and requires the sample to be immersed in a suitable volatile solvent. Any flammable liquid present is absorbed by the solvent, which is then removed and, if necessary, concentrated through evaporation. The technique is at least as sensitive as distillation and is relatively simple. The major disadvantage of solvent extraction is that many pyrolysis products, dyes and other chemicals are also extracted, in some cases presenting major difficulties for the subsequent analysis.

Solvent extraction can be used for solid or liquid samples. It has particular uses in small samples and in the separation of small amounts of flammable liquid remaining in containers.

Clean-up

Clean-up refers to a process of cleaning a sample or extract to prepare it for analysis. Direct injection of headspace vapor and solid-phase microextraction techniques, by their nature, render clean-up unnecessary or impractical. However, in extraction techniques, which result in the preparation of a small liquid aliquot for analysis, clean-up can be useful and occasionally essential.

A simple clean-up procedure is filtration. This can be used to remove suspended solids from a neat liquid sample, or those resulting from a solvent extraction procedure. More complex clean-up techniques involve processes such as acid stripping.

Acid stripping is used to remove oxygenated or nitrogenated compounds that arise from combustion of synthetic materials. These compounds can interfere with analysis of flammable and combustible liquids and addition of a small amount of strong mineral acid removes these compounds by extraction into the acid. The immiscible hydrocarbon layer can be analyzed in the normal manner.

Chromatographic techniques, such as solid-phase extraction and thin layer chromatography, can be useful for clean-up of larger samples but are difficult to adapt to smaller quantities of flammable liquid. With solid-phase extraction, for example, the adsorption extract can be separated on suitable columns into aliphatic and aromatic hydrocarbon fractions. These fractions can be analyzed separately, thus minimizing the interference caused by pyrolysis products.

Any procedure conducted on a sample increases the possibility that the sample will be contaminated or otherwise adversely affected. Although this can be monitored by appropriate use of control samples, the most effective method of reducing clean-up contamination problems is by the minimal application of clean-up procedures. There is a further advantage in minimal clean-up; the overall method remains relatively simple. Addition of a clean-up step is a significant increment in the complexity of the analysis, and the use of that step is only justified by a significant increase in the information recovered.

Analysis

Essentially the same analytical procedures are adopted for the various classes of flammable and combustible liquids. The liquids can be derived from petroleum, and indeed, the most common flammable or combustible liquids detected at fire scenes are petrol (gasoline), kerosene, automotive diesel fuel (distillate) and mineral turpentine. Despite the obvious differences in volatility, the same extraction, clean-up and analysis schemes serve for identification

of all these products, and for the identification of similar products such as white spirit and various naphthas. However, there are other flammable and combustible products, not of petroleum origin, such as pure or wood turpentine, essential oils, ethanol (methylated spirit), natural oils and waxes.

Gas chromatography (GC) GC is the principal tool for the identification of these products. Most petroleum products are composed of many chemical components. Because there are relatively strict specifications which apply to petroleum products, in each different product the components will be present in a ratio which is characteristic of that product. Analysis by gas chromatography separates the individual components, presents the components according to volatility and identifies the amount of each component. This produces a 'fingerprint' of the product that can be compared with known standards for identification purposes.

Several liquids manufactured from petroleum, and many other commercial products (e.g. solvents such as methyl ethyl ketone or methyl isobutyl ketone, ethanol, etc.) have a single component, and it is not possible to identify a product on the basis of the chromatogram retention time alone. Products such as gasoline, kerosene and automotive diesel fuel, and even nonfuel solvents such as mineral turpentine or white spirit, can usually be identified in gas-liquid chromatography by a 'fingerprint'-type comparison of a complex mixture of features. The features considered in any such comparison will include relative retention times, peak heights and peak height ratios. A single component substance obviously does not allow relative peak heights or peak height ratios to be compared, leaving only comparison of retention times. It is necessary, in these cases, to adopt a procedure that will allow the identity to be confirmed. This may consist of repeating the analysis with a 'spiked' sample, to determine whether the sample and standard co-elute, that is, present a single peak. More commonly, the analysis is completed on a gas chromatograph/mass spectrometer.

Modern laboratories use a technique known as capillary gas chromatography, with a specialized detection system called a flame ionization detector (FID), to easily detect microgram and even nanogram amounts of hydrocarbon products.

Gas chromatography/mass spectrometry (GC/MS) GC/MS uses the same component separation procedure as gas chromatography but, as the various components are detected, they are individually analyzed and compared with a library of standards to positively identify the individual components. In the case of a single component product, this may be regarded as a positive identification. With specialty products such as solvents or thinners, it may still be necessary to run comparison standards.

Specialty products are a major challenge in terms of brand identification. Most product types are similar, and examination of the components that make up the product may enable it to be identified in a generic sense. Positive brand identification is much more difficult to achieve, as it requires elimination of other possibilities. The identification of such a product is likely then to be as '... a toluene/xylene based thinner ...' rather than as ... Acme VM&P Naphtha ...'. However, for evidentiary purposes, it may be appropriate to also describe the toluene/xylene based thinner as '... consistent with the composition of Acme VM&P Naphtha ...'.

Techniques in GC/MS, such as selected ion monitoring (SIM) allow identification to be made on the basis of the ratio of selected ions characteristic of the product/s of interest. This can reduce, but not eliminate, the effects of contamination from various sources, including pyrolysis products. SIM reduces the complexity of the comparison to a manageable level, although there are now pattern matching and library programs available that can considerably alleviate the difficulty of identifying an unknown.

Infra-red spectrometry/Fourier transform infrared spectrometry (IR/FTIR) IR/FTIR is an analytical instrumental technique used widely in chemical analysis to identify the chemical 'fingerprint'. An infrared spectrum is obtained that can be particularly useful for the identification of plastics and synthetic materials. The spectrum derived from the infrared analysis reflects the chemical bonding of the molecule, providing some information in relation to the structure of the unknown substance.

The particular advantage of FTIR is that extremely small samples can be analyzed rapidly and accurately. As with GC and GC/MS, there are computer-based pattern matching programs to match sample spectra to library spectra that considerably reduce the difficulty of interpretation of complex spectra. In using spectral libraries, care must be taken not to limit the search to the library contents, because even the largest reference libraries do not cover every possible product.

Because infrared spectrometry analyzes all the chemical groups present in the sample, only samples that are essentially unadulterated can be identified using IR/FTIR.

Microscopy/scanning electron microscopy These techniques are applied in the analysis of debris and

residues from fire and explosion scenes. An unidentified object or unknown material can be identified by examination using a relatively low power microscope. This may permit recognition of components of an incendiary device, of details on paper or material used as fuel, or a physical match of evidence found at the scene to an item found at a suspect's house. Particles or crystals may be selected for more detailed examination, under a higher power microscope or even a scanning electron microscope (SEM).

The SEM, with the capability of magnification to 300 000 ×, can be used to study the morphology of materials and for X-ray analysis. The SEM can provide a rapid elemental analysis of microscopic samples, detailing the amount and distribution of the medium to heavier elements. The actual shape or surface characteristics of the material being examined may be of significance in some cases but, more commonly, the microscope is used to locate materials requiring analysis. This can provide a starting point for further investigations, possibly using additional instrumental techniques. The presence of barium, for example, may suggest the presence of sparkler composition, whereas the presence of strontium may suggest the presence of pyrotechnic material from a red handheld flare. In each case, further analysis would be conducted, such as ion chromatography to identify the nitrate or other oxidizing agent present.

Electrical Components and Electrical Wiring

The diagnosis of electrical sources of ignition is one of the most challenging aspects of fire investigation. The scene examination process may result in the identification of one or more possible sources of ignition, which may or may not be electrical. A laboratory examination might be required to determine whether, say, an appliance was a possible source of ignition, whether a timing device was operating a light switch or an incendiary device, etc.

Examination of individual electronic components with respect to their intrinsic properties, for example, measurement of the exact resistance or capacitance, is often beyond the capabilities of the forensic science laboratory. This is a specialized examination, requiring electronic engineering expertise, which is likely to be found only in universities or major telecommunications companies. The same may be said for integrated circuits and programmable microprocessors. A forensic scientist might identify them as such, but to determine exactly how the microprocessor operates is almost certainly a task for an electronic engineer.

Examinations requiring laboratory expertise are often based on the application of simple physical science. For example, a burnt and melted mixture of components and plastic might be examined using infrared or ultraviolet light, or X-rays. By means of X-ray examination in particular, it is sometimes possible to trace the circuit outlines clearly and identify the construction and means of operation of the components. Similarly, with individual components, the laboratory examiner is often able to identify a clock, wires, a battery and a short length of resistance wire. Obviously, these are the components required for a crude time-delay ignition mechanism. It is possible to calculate the current that can be drawn, and the heat output of the resistance wire, to determine whether this is a functional system.

Laboratory examinations can enable appliances such as power boards or double adaptors to be identified. Careful cleaning, removal of melted material and microscopic examination can enable determination of the position of switches, or whether appliance pins were present in power points, or whether a thermal cut-out switch has operated. These examinations are technically straightforward, albeit they require some practical experience and theoretical knowledge. Clearly though, they can be of great significance since they can result in a definite answer to a potentially crucial question, such as whether a suspect appliance was connected to power or turned on.

Formulating and addressing these basic questions, and interpreting the results of consulting electronic engineers on more complex questions, is an appropriate role for a forensic science laboratory.

There are many proponents of theories relating to electrical arcing. Arcing is a well-known and understood phenomenon, which occurs under specified conditions and which can cause ignition of combustible material. When evidence of arcing is produced, there is the difficulty of determining whether it occurred prior to or as a result of the fire. Any microscopic examination of electrical wiring to determine whether arcing has occurred should be combined with scene observations to determine whether the origin of the arcing is consistent with the apparent point of origin of the fire.

Related theories propose that the chemical composition of copper wires, that is, the relative oxygen and/or chlorine content of electrical wires, as determined by Auger electron spectroscopy or scanning electron microscopy, can be used to ascertain the atmospheric conditions to which the melting wire was subjected. The basis of this theory is that high oxygen and/or low chlorine levels are consistent with arcing prior to the fire; low oxygen and/or high chlorine levels are consistent with arcing caused by the fire. Indeed, in controlled laboratory conditions, these theories can hold true.

However, in the wildly fluctuating atmospheric conditions existing in structure fires, and the many other factors such as varying temperatures, exposure to firefighting water/chemicals, residues from household/industrial chemicals, variations in original wire and pre-existing damage from age or exposure, simple laboratory principles are found to be inadequate. These theories are not transferable from the static laboratory environment to the dynamic environment of an actual fire scene and this type of scene/laboratory examination is neither well established nor practical.

Machinery

The principles for the examination of machinery in the forensic science laboratory are essentially the same as for examination of electrical or electronic appliances or components. The initial role of the forensic scientist is to formulate the questions that need to be addressed in the examination of the machinery. Some of these may be within the province of the forensic scientist, such as whether the machinery can cause sparks in some manner, what is the normal operating temperature and does partially blocking the exhaust markedly increase manifold temperatures? All these are questions that can be resolved by careful measurement and/or observation.

With machinery which is alleged to have malfunctioned, and which has been operated incorrectly or which has been severely damaged, it may be necessary to consult with the manufacturer or some more specialized agency, for example universities and consulting engineers. More complex questions, or questions relating to efficiency, ergonomics or other design factors may be addressed by specialist investigators. This may be under the supervision of a forensic scientist or police officer for continuity purposes, as a consultant may not be aware of the requirements for security and continuity of exhibits.

Summary

This article has briefly reviewed the wide range of scientific expertise and techniques necessary to support the fire scene investigator in a scientific examination and interpretation of a fire scene. However, it must be remembered that the quality of the fire-scene interpretation and of the samples taken will be directly responsible for the subsequent quality and usefulness of the laboratory results. Furthermore, these laboratory results will only rarely provide any proof of the fire cause. They will, in fact, provide supporting evidence to the scene interpretation so that possible fire causes can be identified and, perhaps, eliminated or upgraded to probable causes.

See also: **Analytical Techniques**: Separation Techniques; Mass Spectrometry. **Crime-scene Investigation and Examination**: Recording; Collection and Chain of Evidence; Recovery of Human Remains; Packaging; Preservation; Contamination.

Further Reading

Almirall JR, Bruna J and Furton KG (1996) The recovery of accelerants in aqueous samples from fire debris using solid-phase microextraction. *Science and Justice* **36**:283–288.

ASTM Standard E1385-95, *Standard Practice for Separation and Concentration of Flammable or Combustible Flammable Liquid Residues from Fire Debris Samples by Steam Distillation*.

ASTM Standard E1386-95. *Standard Practice for Separation and Concentration of Combustible Liquid Residues from Fire Debris Samples by Solvent Extraction*.

ASTM Standard E1387-95. *Standard Test Methods for Flammable or Combustible Liquid Residues in Extracts from Samples of Fire Debris by Gas Chromatography*.

ASTM Standard E1388-95. *Standard Practice for Sampling of Headspace Vapours from Fire Debris Samples*.

ASTM Standard E1389-95. *Standard Practice for Cleanup of Fire Debris Sample Extracts by Acid Stripping*.

ASTM Standard E1412-95. *Standard Practice for Separation and Concentration of Flammable or Combustible Liquid Residues from Fire Debris Samples by Passive Headspace Concentration*.

ASTM Standard E1413-95. *Standard Practice for Separation and Concentration of Ignitable Liquid Residues from Fire Debris Samples by Dynamic Headspace Concentration*.

Brackett JW (1955) Separation of flammable material of petroleum origin from evidence submitted in cases involving fires and suspected arson. *Journal of Criminal Law, Criminology and Police Science* **46**:554–561.

Gilbert MW (1998) The use of individual extracted ion profiles versus summed ion profiles in fire debris analysis. *Journal of Forensic Science* **43**:871–876.

Juhala JA (1979) Determination of fire debris vapours using an acid stripping procedure with subsequent gas chromatographic and gas chromatography mass spectrometry analysis. *Arson Analysis Newsletter* 3(4):1–19.

Karkkainen M, Ilkka S and Himberg K (1994) Detection of trace levels of gasoline in arson cases by gas chromatography–mass spectrometry with an automatic on-line thermal desorber. *Journal of Forensic Science* **39**:186–193.

Lennard C (1995) *Fire (determination of causes) a review 1992 to 1995*. 11th Interpol Forensic Science Symposium, Lyon, France.

Mickiff C (1978) Recovering accelerants from debris. *Arson Analysis Newsletter* **2**(6):8–20.

Twibell JD, Home JM and Smalldon KW (1980) A comparison of the relative sensitivities of the adsorption wire and other methods for the detection of accelerant residues in fire debris. *HOCRE Report* no. 368.

Waters LV and Palmer LA (1993) Multiple analysis of fire debris samples using passive headspace concentration. *Journal of Forensic Science* 38:165–183.

Woycheshin S and De Han J (1978) An evaluation of some distillation techniques. *Arson Analysis Newsletter* 2(5):1–16.

Yip IH and Clair EG (1976) A rapid analysis of accelerants in fire debris. *Canadian Forensic Science Society Journal* 9(2):75–80.

Physics/Thermodynamics

J H C Martin and **R S Pepler**, Institut de Police Scientifique et de Criminologie, Université de Lausanne, Lausanne, Switzerland

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0525

Introduction and Overview

Within the various domains of criminalistics, the investigator compares traces found at the scene of a crime with those recovered from a person, tool or place. By contrast, the work of the fire investigator is not purely comparative in nature; extensive practical experience of the examination of as many fire scenes as possible combined with a sound scientific knowledge, especially in the areas of physics and thermodynamics, is paramount to the success of the investigations. These two criteria enable him or her to identify the origin of the fire and determine its cause despite having to work on sites that have been destroyed by the fire and further disrupted during the extinguishing process.

This article explores: the relevant background areas of physical thermodynamics; the role of thermodynamics in fire investigation; fire ignition and propagation; thermodynamic classification of ignition sources; and the phenomena of smouldering and flaming combustion.

Physical Thermodynamics: the Relevant Background

Before considering the role of thermodynamics in fire investigation, it is essential to remember that the fundamental subject under scrutiny is concerned with the transformation of a macroscopic system which is dependent on one of the basic elements of physics: temperature. It is also important to understand that an intensive thermodynamic variable is one which is dependent on the amount of a chemical substance in the system, whereas an extensive variable does not.

Physical systems

A physical system is defined as that part of the universe under the influence of thermodynamics which has been identified for study. It can be:

- isolated: no exchange of heat, work or matter with the surroundings is possible. The system is said to be ideal and can be assumed in various situations;
- closed: exchange of heat and work, but not matter, is possible with the surroundings;
- open: exchange of heat, work, and matter is possible with the surroundings.

Thermodynamic principles

Two of the fundamental laws of thermodynamics can be described in the following manner:

The first law of thermodynamics All systems possess an internal energy U, which is a state variable, signifying that U is independent of the history of the system. When a system undergoes a change from state 1 to state 2, the difference in thermodynamic energy, ΔU is:

$$\Delta U = U_2 - U_1$$

While respecting the principle of the conservation of energy, the ΔU of a closed system can be expressed as:

$$\Delta U = \Delta Q + \Delta W$$

where ΔQ and ΔW are, respectively, the applied heat and work.

For an isolated system U is constant; in keeping with the conservation of energy, U_1 is therefore equal to U_2. There is no energy change between the system and its surroundings therefore:

$$\Delta Q = \Delta W = 0$$

The second law of thermodynamics This law concerns change and entropy. Entropy, S, is an extensive state variable which denotes the degree of disorder of the system. During reversible changes, the system and surroundings are in constant equilibrium. When considering such a change from a state 1 to a state 2, the entropy of a system is:

$$S = \int dQ/T$$

irrespective of the pathway taken.

However, energy dissipated in the form of heat is not an ideal process; it is spontaneous and irreversible

which consequently means that the entropy of such a system increases.

The Role of Thermodynamics in Fire Investigation

An understanding of fundamental thermodynamics is important in fire investigation in considering possible ignition processes. The most important factors are identified below.

Temperature

Temperature is an intensive state variable which can be measured for any given system by the use of a thermometer. The Celsius, or centigrade, scale of temperature is governed by the triple point and the boiling point of water. At 0°C, under standard pressure, water exists simultaneously in solid, liquid and gaseous states; at 100°C only the existence of the liquid and gaseous states is possible. There are two other temperature scales which are extensively used: Kelvin (K) and Fahrenheit (F). The conversion formulae between the three scales are as follows:

$$0^\circ\mathrm{C} = 273.15\ \mathrm{K} = 32^\circ\mathrm{F}$$

$$\mathrm{T\ (K)} = \mathrm{T\ (^\circ C)} + 273.15$$

$$\mathrm{T\ (^\circ F)} = 9/5\ \mathrm{T\ (^\circ C)} + 32$$

At any given atmospheric pressure there are certain temperature values, which are specific to a particular substance, that indicate whether the ignition of this substance is possible.

- Flash point: the lowest temperature at which a sufficient quantity of vapor is produced above a solid or liquid which can be ignited by a flame, spark or localized heat source. At this temperature combustion does not endure;
- Point of inflammation: the lowest temperature at which the production of flammable vapors above a solid or liquid is such that a flame, spark or other localized heat source will cause their ignition; combustion and its accompanying flames are sustained;
- Point of auto-ignition: the lowest temperature at which a solid or liquid will spontaneously ignite without the presence of an additional heat source (flame, spark or other).

Heat

Heat is a form of energy which has a number of effects on a system. It can be generated by the transformation of energy, be it mechanical, electrical, chemical or other, into calorific energy. A temperature gradient exists between two systems at different temperatures, along which calorific energy (heat) is transferred in order to reach a state of thermal equilibrium.

The fundamental unit of calorific energy is the Joule (J); however, a more widely recognized terminology is the calorie (cal) or kilocalorie (kcal). These units are defined as the quantity of heat required to raise 1 g, or 1 kg respectively, of water from 14.5°C to 15.5°C.

$$1\ \mathrm{kcal} \equiv 10^3\ \mathrm{J} \qquad 1\ \mathrm{cal} \equiv 4.185\ \mathrm{J}$$

Physical properties

Heat capacity, C, is the heat transferred to a body in order to raise its temperature by one degree (K):

$$C = \frac{\mathrm{d}Q}{\mathrm{d}T} \qquad \text{units: J K}^{-1}$$

Specific heat capacity, c, is the heat capacity of 1 kg of a substance which, in general, is dependent on temperature:

$$c = \frac{1}{M}\frac{\mathrm{d}Q}{\mathrm{d}T} \qquad \text{units: J kg}^{-1}\ \text{K}^{-1}$$

One of the fundamental laws of physical chemistry is the Ideal Gas Law:

$$pV = n\boldsymbol{R}T$$

where p = pressure, in Pa (1 atm = 101325 Pa) V = volume, in m^3; n = moles; T = temperature, in K; R = gas constant = 8.315 J K^{-1} mol^{-1}.

When an ideal gas makes the transition from state 1 to state 2 the following combined gas law applies:

$$\frac{p_1V_1}{T_1} = \frac{p_2V_2}{T_2}$$

This expression can be used for real gases on the understanding that the data obtained will only be an indication of and not the exact values desired.

Fire: Ignition and Propagation

Fire is an oxidation–reduction (redox) reaction accompanied by the production of heat and light. The two chemical requirements are an oxidizing agent (or combustive agent) – usually the oxygen in air – and a reducing agent (or fuel), which must be in a gaseous state in order to mix with the air. An activation energy, always in the form of heat, is required to ignite the gaseous mixture and combustion is maintained by the calorific energy produced during the exothermic reaction.

Activation energy

The limit of flammability of a gas (or gaseous mixture) is expressed as a percentage which indicates the

proportion of gas present in the air. The ignition of a homogeneous mixture of air and a gas is possible when the lower limit of flammability has been attained; however, ignition is impossible above the upper limit of flammability as the air is too saturated. Understanding and establishing the activation energy that is required for ignition of a gas/vapor (other than autocombustion) is fundamental to fire and explosion investigations. The scene examination is often complex as several calorific energy producing systems can be present, of which some could be weak heat producers and others situated some distance from the combustible material. In any situation, the means of ignition has to originate from one of these sources. In order to eliminate each ignition possibility until left with the most probable, a second factor must be taken into consideration: the thermal or calorific power.

Thermal power and reaction temperature

The thermal or calorific power of a system is the calorific energy produced over a period of time. The investigator can only estimate the thermal power which is at the origin of and aids the development of a fire. The relationship between the thermal power of the heat source, W_T, and that dissipated in the surrounding environment, w_T, constitutes the thermodynamic foundation of the whole investigation. This can be seen from the following equation derived from the Arrhenius expression:

$$W_T = \Delta H_c n A \exp(-E_a/RT)$$

where: ΔH_c = enthalpy of combustion; A = pre-exponential factor, E_a = activation energy and:

$$w_T = hA\Delta T$$

where: h = convective heat transfer coefficient; A = surface area.

The thermal power of a heat source is exponentially dependent on temperature, and the only circumstances when a body will not emit heat is at absolute zero temperature (0 K); $W = 0$ if, and only if, $T = 0$.

This observation is important in that any system not at $T = 0$ K will produce calorific energy. If the energy dispersion into the surrounding environment is restricted, the system's temperature will inevitably increase. Ignition frequently occurs where the temperature of a combustible material has been augmented to the point of autocombustion.

The possibility of a given heat source (spark, flame, fermentation, etc.) igniting a specific material (wood, insulation, paper, etc.) in a certain environment (quasi-isolated, closed or open) must always be considered before the source can be eliminated from the investigation. However, calculations of W_T and w_T for real systems are impractical and inexact, except for certain types of heating; instead, the essential factor that should be taken into consideration when examining possible ignition sources is the influence of W_T and w_T on the system. Therefore, the first step of a fire/explosion investigation is to locate the place where the activation energy was produced; in other words, to determine the foyer, or seat, of the fire/explosion. Once a potential heat source has been identified, the implications of the thermal power produced and the heat dissipated in its immediate vicinity should be estimated. This enables the investigator to verify whether or not the proposed source is a plausible source of ignition of the material situated at the (presumed) foyer.

Combustion speed

The phenomena of smouldering fires, flaming fires and explosions differ only in their rate of combustion.

- Smoldering is a flameless combustion involving the slow oxidation of a porous combustible material, which occurs in an environment where oxygen can easily penetrate yet is sufficiently isolated so that heat dissipation to the surroundings is minimal.
- The speed of combustion of a flaming fire can vary enormously. The chemical reaction occurs without a perceptible pressure increase within the immediate vicinity.
- An explosion is an extremely rapid combustion. When occurring in a confined environment it is accompanied by a considerable augmentation of pressure. The reactants are present within their limits of flammability and are usually in a gaseous state; however, very fine droplets of liquid, such as those produced by an aerosol, or fine powders/dust can also combust explosively. After ignition, the entirety of the explosive material undergoes the exothermic reaction of combustion almost instantaneously; the ambient air subsequently expands very quickly resulting in a pressure increase in the immediate environment.

Thermodynamic Classification of Ignition Sources

Fires can be classified according to their ignition source, whether due to a spark, naked flame or localized heat source.

Heating

Heating is a transfer of calorific energy to a physicochemical system without the presence of a spark or flame. If the system generates heat itself to augment its temperature to the point of inflammation and

subsequently to autoignition this is known as spontaneous combustion. With the exception of this latter form of heating, heat is transferred to the material via the methods outlined below.

Conduction This is the mode of heat transfer without displacement of matter, due to electronic movement within a solid. Where two faces of a solid are at different temperatures, T_1 and T_2, the heat flux, q', through the material is:

$$q' = \frac{k}{x}(T_1 - T_2) \qquad T_1 > T_2$$

where: x = thickness of material; k = thermal conductivity.

If a steady state heat flux traverses two (or more) layers of material:

$$q'_1 = q'_2 = q'_{1,2}$$

$$q'_{1,2} = \frac{T_1 - T_2}{x_1/k_1 + x_2/k_2}$$

Convection Convection occurs in liquids and gases and is a transfer of heat via mechanical movement due to changes in density. This process can, however, transfer heat to a solid. For example, if a solid surface of area A is in contact with a liquid/gas at temperature T, the heat flux, q'', transferred to the surface is:

$$q'' = hAdT \qquad \text{units: Watts}$$

where: h = convective heat transfer coefficient; A = surface area.

Radiation Electromagnetic radiation in the visible and infra-red regions is radiated from all bodies with a temperature above absolute zero. When a system is in thermal equilibrium its components emit and absorb radiated heat at the same rate. Although this radiation is rarely quantified it is important to note that:

- radiation is the prevalent mode of heat transfer between flames and combustible material; it is therefore the principal cause of vaporization and consequently the main means of the ignition of fresh material;
- before suspecting human intervention after a large fire where several foyers have been identified, it is essential to verify that the separate ignition sites are not a consequence of thermal radiation. Fire may have spread in this manner from one combustible material to another despite the lack of an apparent physical link.

When a body has two identical surface areas, A, respectively at temperatures T_1 and T_2, a heat flux, f, flows from surface 1 to surface 2 if $T_1 > T_2$:

$$f = a_1 a_2 \sigma A({T_1}^4 - {T_2}^4)$$

Where: a = heat absorption or reflection factor; σ = Boltzmann constant.

Heating appliances When a heating appliance is identified as being the origin of a fire, any of the three modes of heat transport can constitute the transfer of calorific energy to the combustible material. As the correct use of these devices is for the generation of heat, the cause of the fire has to be a result of an inappropriate usage or a malfunction of the apparatus.

Electrical sources Fires of electrical origin are often complex to investigate and difficult to prove. Heat is produced when an electrical current, I, passes through a conductor with resistance, R, over a time period, t, according to Joule's Law:

$$\Delta E_{\text{electric}} = I^2Rt \qquad \text{units: } A^2 \cdot \Omega \cdot \text{sec} = \text{J}$$

The electrical power is therefore:

$$W = I^2R \qquad \text{units: } \Omega A^2 = \text{W}$$

Friction The generation of heat by the friction between two objects is proportional to the speed at which the two bodies move against each other and their respective weights. The nature of the surfaces in contact, which is indicated by the friction coefficient, will also have a bearing on the quantity of calorific energy produced; those with a high friction coefficient generating more heat.

Sparks **Material sparks** Material sparks are produced when solid particles with a mass of milligram magnitude are emitted into an open system existing at high temperature. These tiny fragments of incandescent or burning matter can constitute the origin of a fire if:

- when smoldering, the distance between the source and the combustible material is very short; due to their miniscule mass the particles lose heat rapidly;
- the spark enters into contact with a highly flammable material or one with a small heat capacity in which a smoldering fire can manifest.

A spark of mass m, and heat capacity c, emitted with a temperature T_2, into an environment at temperature T_1, produces heat, Q, according to:

$$Q = mc(T_2 - T_1)$$

For example, when grinding a metal, the metallic fragments produced are of milligram mass, which generate an approximate heat of $Q \approx 1.3$ J; this is of

large enough magnitude to ignite a flammable gas/vapor. Another example is the fusion of a metal/metal oxide when welding; the resulting droplets have a diameter of 2–5 mm and generate heat between the values of: $37 < Q < 578$ J. The first value is more than sufficient for the ignition of a gaseous mixture, whereas the second can initiate the combustion in a nearby combustible material such as plastic isolation or wood, provided that it is within the boundaries of the quasi-isolated system.

Electrical sparks and arcs In the space separating two charged objects, there is an electrical field. If an avalanche of electrons occurs due to this field it is visible (to the naked eye) as a spark. Because of its short lifetime, an electric spark produces a relatively small amount of heat; nevertheless, the millijoules of energy generated can ignite a homogeneous air/flammable gas mixture resulting in an explosion.

The difference between an electrical arc and spark is the duration of the electrical discharge: that of an arc being markedly longer than a spark, which can be said to be instantaneous. Logically, it follows that the energy produced by an arc is of greater magnitude and can subsequently inflame a combustible material rather than uniquely gaseous mixtures as for a spark.

The above phenomena will occur when in the following conditions.

- A defective contact (of a switch, relay unit, etc.) in an electric circuit, with or without a coil, will result in an increase in resistance and will thus generate heat. This default can also occur with the use of adapters and extension cables.
- In the presence of an electrolyte (water, humid air, etc.), a defective insulation of a cable will enable the passage of an electric current between two conducting wires at different electric potentials. An augmentation in resistance will subsequently generate calorific energy at a precise point which can be sufficient to ignite the insulation or a combustible material in close proximity.

Smoldering

In a smoldering solid the heat conduction is difficult to calculate, as are the speed and propagation of combustion. This type of fire obviously still requires an ignition source, however the thermal power of the source can vary enormously. For example:

- after most fires, smoldering is the natural progression of flaming combustion. This latter form of combustion can be reinitiated when a great quantity of smoldering material is remnant after a large fire. This eventuality only applies to one aspect of fire investigation: that where flames are rekindled from smoldering remains, even after extinction by the fire brigade;
- when the heat source is weak, smoldering combustion commences at the surface of the material. In this case the only conceivable progression of combustion is to the interior of the material, otherwise self-extinction will occur.

Smoldering combustion is self-maintained if an equilibrium is established between the thermal power of the exothermic (combustion) reaction and that dissipated into the surroundings:

$$\Delta W = W_T - w_T$$

If the difference, ΔW, between the former thermal power values is small, the fire will neither self-extinguish nor reach an open flame state; the smoldering combustion will simply continue to burn fresh material with a random progression that is due to the inhomogeneous environment. In order to maintain a constant value of ΔW, or one that varies within a very narrow range, the oxygen supply must be adequate; nevertheless, it must be noted that even in an oxygen-deficient environment (proportion of O_2 in the air $<16\%$), it is possible that this type of combustion will be sustained.

Smoldering material will burst into flames if there is sufficient heat build-up and enough oxygen available. For example, smoldering combustion will progress to the flaming stage if the burning material enters into contact with air. A pertinent example of this is when the exterior surface temperature of a material is increased by heat transferred from internal smoldering combustion; when the surface, which is in contact with the ambient air, reaches the point of inflammation flaming combustion will result and will spread at a rapid rate.

Flames

If a combustible material is exposed to prolonged contact with a flame its temperature will eventually reach the point of inflammation and ignition will occur. The resulting flames will subsequently raise the temperature of the surrounding material which will also attain its inflammation point and start to burn. The propagation of fire by direct flame contact is a natural progression which does not require further explanation.

Hot gases produced by burning material are propelled upwards. Due to their inferior density the buoyancy forces within the smoke plume cause the hot gases to rise with a speed that is dependent on the temperature of the fire. At high temperatures, for

example directly above the flames, the gases are dispersed vertically very rapidly and have little sideways diffusion into the ambient air. This effect diminishes as the ambient temperature decreases, thus the gases have greater horizontal convection the further they are from the flames. They can therefore be seen to rise in the form of a three-dimensional 'V' (an inverted cone). This plume of hot gases transports soot particles which are deposited on vertical nonflammable supports; the characteristic triangular pattern of these deposits can often be used to locate the seat of a fire.

The smoke resulting from a fire contains combustion products and small unburned residues which are, by definition, nonoxidized vapors that are transported by the diffusion of the other hot gases. These vapors can accumulate in enclosed spaces and spontaneously ignite, under the correct conditions, even at considerable distances from the original foyer.

Fire spread can often be deduced with relative ease if the initial heat source has been identified and there is evidence of the presence of combustible material throughout the entire burned area. However, when justifying a source of ignition, problems can arise when the heat source is found to be separated from the combustible material by nonflammable material such as walls, floors and ceilings.

With an open fire, the vertical flame movement and the air supply are practically limitless. At the exterior of a stone or brick building, a fire situated at ground level can extend to roof height. If the roof is constructed from combustible material such as wood or thatch, it will eventually catch fire (due to the prolonged contact with the flames and hot gases) even if separated from the seat of the fire by several floors of nonflammable material.

If the fire is restricted by a wall or other incombustible barrier, the supply of air may well be limited. The flames will consequently spread towards the direction of the air source. When vertically extending flames are hindered by the ceiling of a room, the plume of hot gases is redirected along its surface. The horizontal spread can be of considerable distance as the area directly below the ceiling is at a highly elevated temperature, thus sustaining the capacity of the gases to ignite fresh combustible material. Therefore:

- if the entrainment of air and the diffusion of flammable gases/vapors in the vertical part of the flame is sufficient, there will be limited flame spread: combustion will occur at ceiling level;
- with a low ceiling or a vapor-rich smoke plume, the hot gases will spread out beneath the ceiling and the flames can spread to material situated some distance from the original fire seat.

It must be remembered that a corridor, ventilation shaft or other form of passageway can aid the propagation of fire between rooms, apartments or even separate buildings.

Conclusion

This article does not attempt to list each and every possible heat source capable of initiating a fire. Instead it indicates the most significant sources and analyzes them thermodynamically, consequently highlighting the role of thermodynamics and that of fundamental physics in verifying the conclusions of the fire investigator. Nevertheless, it must be noted that numerical values are rarely calculated when considering the various exchanges of heat or the laws of thermodynamics, as the fires under investigation were not performed in the laboratory under controlled conditions. The equations and mathematics are simply used to give an indication of the magnitude of the process in question. The work of the fire investigator is therefore to identify the origin of a fire and to report its cause by exploiting an extensive practical experience and intricate knowledge of the laws of physics and chemistry.

See also: **Crime-scene Investigation and Examination**: Major Incident Scene Management. **Criminalistics**. **Fire Investigation**: Types of Fire; Chemistry of Fire; Fire-scene Patterns. **Health and Safety**: Including Risk Assessment.

Further Reading

Atkins PW (1990) *Physical Chemistry*, 6th edn. Oxford: Oxford University Press.

Bush LS and McLaughlin J (1996) *Introduction to Fire Science*. Fire Engineering. Tulsa, USA: Penn Well.

Chaussin C, Hilly G and Barralis J (1969) *Chaleur et Thermodynamique*. Paris: Dunod.

Clavin P and Linau A (1985) *Theory of Gaseous Combustion*. NATO ASI SER; SER B **116**:291–338.

DeHaan J (1991) *Kirk's Fire Investigation*, 3rd edn. Brady: Prentice Hall.

Drysdale D (1986) *An Introduction to Fire Dynamics*, 2nd edn. Chichester: John Wiley & Sons.

Noon R (1992) *Introduction to Forensic Engineering*. Boca Raton: CRC Press Ltd.

Phillipps CC and McFadden DA (1986) *Investigating the Fireground*. Fire Engineering. Tulsa, USA: Penn Well.

Schultz N (1985) *Fire and Flammability Handbook*. New York: Van Nostrand.

Yallop HJ and Kind SS (1980) *Explosion Investigation*. Harrogate: The Forensic Science Society and Scottish Academic Press.

Yereance RA (1989) *Electrical Fire Analysis*. Springfield, IL: Charles C. Thomas.

Types of Fire

P J Thatcher, Forensic Services, Northern Territory Police, Fire and Emergency Services, Darwin, Northern Territory, Australia
J D Kelleher, Fire and Explosion Investigation Section, Victoria Forensic Science Centre, Melbourne, Victoria, Australia

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0524

Introduction

The destructive and dangerous properties of fire have led to the allocation of enormous economic resources to fire detection and suppression, fire investigation techniques and materials and product safety research. Also, the aftermath of fire often requires additional economic resourcing by insurance companies and, in times of natural fire-related disasters, by governments.

As a consequence, the many different professions that have a direct and indirect interest in fire have designed different classification systems for types of fires to reflect their particular interest. For example, when ascertaining eligibility for compensation, it is important to determine whether a fire is the result of:

- negligence
- an accident
- a malicious act

However, when public safety is an issue, it may be important to classify fires according to what was actually burned, for example

- dwellings
- commercial premises
- vehicles
- public property (schools, railway stations, churches, etc.)
- marine vessels
- parks and reserves

When fire investigation is the primary consideration, types of fire can be classified by the source of ignition, i.e. the fire cause: e.g.

- electrical
- machinery
- autoignition
- weather (lightning)
- direct ignition

It should be noted that in this final classification scheme there is no mention of intent, even in the case of a match or a cigarette lighter. This is because the effect of ignition is identical regardless of degree of intent, and a purely scientific examination for fire cause can not usually prove or refute intent. However, the presence of other evidence, such as separate and unconnected fires and/or a time-delay ignition device, will provide indisputable evidence of intent.

In this article, types of fire are classified according to the source of ignition, although on numerous occasions reference will be made to the ignition and burning characteristics of specific objects such as machinery, motor vehicles, etc. in order to illustrate a particular feature of fire.

This is therefore a rather academic analysis of fire causes. From a practical viewpoint, the various features of the development of fires must also be considered. This is made necessary by the fact that fire investigators must interpret any surviving indicators of fire development before nominating the fire origin and the ignition source.

Sources of Ignition

The fundamental types of fire according to their cause of ignition are described below.

Electrical fires

The most misunderstood of fire causes are those related to electrical origins, and this is the principal reason why some unreliable fire-cause statistics have been compiled over past years.

If electrical appliances and electricity supply are present at a fire scene, any diagnosis of an electrical cause should only be made after the following factors have been carefully considered.

Electrical energy represents a source of ignition. The simplest example is a filament-type gas stove lighter, where electrical energy heats a wire filament until it reaches the temperature necessary to ignite a gas–air mixture. Apart from this obvious example, there are three principal ways in which electricity can be attributed as the cause of ignition of fires.

Electrical sparks

Electrical sparks can ignite flammable or combustible material. This is of most concern when flammable vapor–air mixtures are present, and these mixtures are within the characteristic flammable limits that can be ignited by a relatively low-energy spark.

Significantly, any flammable or combustible material can be ignited by a spark if the energy dissipated by the spark is sufficient to raise the local temperature of the material to the fire point, or flash point, as the case may be. Where hazardous atmospheres are known to exist, intrinsically safe electrical fittings are used to insure that electrical-spark ignition does not occur.

This consideration assumes the presence of a homogenous mixture of gases which, in practice, is unusual and difficult to achieve. It should be noted here that the electrical charge responsible for the spark could be from static electricity, from alternating current electricity or from direct current electricity. The source of the electricity is not important; the significant factor is the thermal energy of the spark.

In the case of solids and liquids, the energy of the spark must be sufficient to raise the local temperature of the substance to the fire or flash point. The energy requirement for most substances is prohibitive; and therefore, fires of this type are likely to occur only in exceptional but important cases.

Electrical wires, cables, connections and components Electrical wiring systems, electrical circuits in appliances and electrical accessories are all designed to avoid overheating from occurring in any section of the circuit. In addition to electrical safety approvals required for individual items, there are more general wiring standards that apply to domestic, commercial and industrial premises. For appliances in good condition in a building that is correctly wired and fused, the chances of an electrical fault resulting in a fire are very small.

Realistically though, not all appliances are in good condition and not all premises are correctly wired. In these cases, electrical fires can be due to:

- damaged or degraded insulation on wires or cables;
- excessive loads on power boards or outlets;
- damaged or dirty electrical contacts;
- inadequate heat dissipation.

In each case, a component or section of a circuit is heated by electrical energy until combustible material nearby is ignited. The energy available in most electrical circuits is many times greater than the energy in a spark, so ignition of combustible material close to an electrical fault is a realistic possibility.

Electrical appliances Electrical appliances can cause fires in a variety of ways. Apart from the circuit or component failures listed above, some possible causes of appliance fires are as follows:

- breakdown of transformer ballast in fluorescent lights;
- overloading of electric motors in washing machines or driers;
- overheating of battery chargers for mobile telephones,
- failure of cooling fans in appliances generating heat.

In all of these cases, the electrical energy supplied to the faulty appliance heats part of the electrical circuit to the point that nearby combustible material is ignited.

Misuse of appliances Despite the increasing use of safety features, including thermal cutout switches, appliances such as electric radiators continue to cause fires when they are placed inappropriately, e.g. too close to curtains. The fire which can result in this way would, in many jurisdictions, be classified as an electrical fire. However, they are more properly classified as 'appliance misuse' because the term 'electrical fire' implies an electrical fault or failure of some type.

Weather

A significant subclass of fires is those caused by natural phenomena. The best documented in this category are fires caused by lightning. These are common and are a recognized problem in forest management. It is not difficult to comprehend that lightning striking a tree, particularly a dead tree, can heat the tree sufficiently for it to ignite. In a summer thunderstorm moving across south eastern Australia, there can be tens, even hundreds, of lightning strikes in forested areas, resulting in many fires.

Lightning strikes on structures occur frequently. Lightning rods or conductors that safely carry the electrical current to earth without damaging the structure protect most tall buildings. Houses and commercial premises are rarely struck by lightning because there are usually many taller and thus more vulnerable targets such as telephone poles or trees in the immediate vicinity. However, a lightning strike on a house or other building could cause a fire or fires in addition to structural damage. There are several factors, which may point to the possibility of lightning as an ignition source:

- a recent thunderstorm with significant electrical activity;
- the affected structure is in an exposed and/or elevated position;
- physical damage to an upper exterior portion of the premises, possibly due to a chimney or aerial.

Apart from lightning, sunlight and rain both have the potential to cause fires in a strictly limited sense. Sunlight can be focused with a lens to cause intense heat and fire and, if the focal point coincides with combustible material, a fire may result. Reports of wildfires (bushfires) being ignited by the sun shining through broken bottles are exaggerated to the point of being untrue. There have, however, been cases of the sun shining through vases or other objects on window ledges and igniting furniture.

Both sun and rain can play roles in the ignition of fires in haystacks and silos through self-heating (often incorrectly referred to as 'spontaneous combustion'). This requires particular conditions of temperature and humidity (see below). Nevertheless, if an agricultural district has been drenched, it may be that there is a sudden outbreak of haystack fires. This might be attributed to the average moisture level of the haystacks increasing to the point where self-heating becomes possible. This type of fire is classified as a 'self-heating fire', but is mentioned here for completeness because of the obvious indirect connection to weather.

Many other climatic conditions can cause fires. Almost any extreme condition can lead indirectly to fires through adverse effects on wiring, cables or machinery.

Machinery

Fires can be caused by machinery. As has been established, for the ignition of flammable or combustible material, that material must be heated to the fire point. Under certain conditions, it is possible for mechanical action to cause a fire by heating combustible material.

An example of ignition by mechanical action is the historic use of the fire bow. The bow is used to rapidly rotate a stick that has one end rotating in a hollow in a wooden plate. Friction causes heating which eventually results in the ignition of finely divided cellulosic material.

A more contemporary example is a fire in a wheel casing. Here, a damaged bearing leads to greatly increased friction and the heat generated can ignite grease and hydraulic fluid in the hub.

Motor vehicles can be ignited as a result of mechanical failure. Many parts, particularly the exhaust manifold, operate at high temperatures. Failure of fuel or hydraulic lines in the engine compartment can result in fuel or hydraulic fluid being sprayed over a high temperature area such as the manifold. If the temperature of the manifold is above the fire point of the fluid, a fire will result.

Fires started by machinery are not necessarily the result of a fault in the machinery itself, but may be due to the method of operation or installation. Many fixed plant items, refrigerators for example, require specified clearances from walls and the floor to allow adequate ventilation for cooling purposes. Failure to comply with these requirements can lead to overheating and result in a fire.

Autoignition

The complex phenomenon of autoignition, or self-heating as it is more properly known, is often incorrectly referred to as 'spontaneous combustion'. There is nothing spontaneous about self-heating, and the principles of chemistry and biochemistry that underlie it are well understood. It is both predictable and avoidable.

Materials which are easily oxidized will, in many instances, have self-heating properties. The spontaneous ignition of haystacks is an example of biological heating. If the hay is in the required moisture range, significant bacterial activity can develop, causing a rise in temperature. A moderate increase in temperature can promote further bacterial growth, with a consequential rise in temperature.

However, the limit to temperature increase as a result of bacterial activity is around 70°C, the upper limit for most bacterial growth.

At or above this temperature, chemical oxidation can further increase the temperature, which can then lead to eventual ignition. Whether the temperature rises to this level depends on the balance between biological/chemical heat production within the stack and heat losses from the bales as a result of conduction and convection. When the heat production exceeds the heat loss, the temperature will rise and the process accelerates. If the heat losses increase (for example, when the external temperature drops significantly), heat loss may exceed heat production and the temperature falls. The probability that the stack will ignite 'spontaneously' is then reduced.

Products containing drying or semidrying oils, such as linseed, cottonseed, sunflower seed and perilla oils are susceptible to self-heating through chemical oxidation and polymerization. This can give rise to the well-documented occurrence of fires in oily rags.

Also, finely divided, readily oxidized materials such as charcoal and coal dust also have the potential for self-heating. This is the cause of many serious industrial accidents in silos, bunkers and processing plants where high levels of dusts are produced.

Direct ignition

Direct ignition is a term that has evolved to classify simple fire causes that do not involve a series of events. The most common cause of direct ignition is the simple application of a flame due to a match or cigarette lighter to a combustible material. This is distinct from, say, a fire due to an electrical cause where some electrical malfunction was the primary event and the fire was a secondary or resulting event.

Direct ignition shares the same basic principles as other types of ignition. A heat source is applied to flammable or combustible material. For that material to ignite, it must be heated to the flash point or fire point. This may take a fraction of a second, as in the

case of a flammable liquid, or several minutes, as in the case of a combustible solid such as wood or plastic.

Match or cigarette lighter Whether a match or a cigarette lighter can cause ignition is dependent on the heat transfer associated with the ignition process. Heat generated by the flame is transferred in part to the material to be ignited. However, if heat is lost by radiation and conduction, only a portion of the heat generated initially serves to increase the temperature of the material.

Heat absorbed by the material to be ignited will raise the temperature towards the flash point or fire point, at a rate that is dependent on the thermal capacity of the material. Material directly exposed to the flame can be oxidized immediately, with heat generated by this reaction contributing towards the overall temperature rise. If the flash point or fire point is reached over sufficient of the material before the heat source, i.e. match or lighter, is exhausted, the generated heat may exceed heat losses, at which point the material is regarded as having ignited.

Once the material has ignited, the combustion reaction is, by definition, a self-sustaining reaction. The development of the fire from this point depends on the nature of the material that has been ignited and on factors such as ventilation. It is not dependent on the method or process of ignition.

Time-delay ignition devices Time-delay ignition devices are occasionally used by arsonists in order to ensure that they are not present when the fire occurs and to allow time for them to establish an alibi in the event that they come under suspicion.

Although fires caused in this way appear to be another 'type of fire', in effect, they can be classified according to the ignition source as can any other fire. The construction of time-delay ignition devices is only restricted by imagination and opportunity. They range from simple devices based on burning candles or cigarettes to sophisticated devices which depend on chemical reactions, electrical appliances, mechanical apparatus (such as mouse-traps), etc. In some instances, items such as clocks and electrical timers have been used to determine the ignition time. Electrical items which can be activated by a telephone call have also been used to initiate ignition.

The most significant feature of fires caused by the ignition of time-delay ignition devices, is the irrefutable evidence of intent which, at least from a purely scientific point of view, is unusual.

Fire Development

The preceding discussion is only of academic interest unless the consequences of ignition sources on subsequent fire development are understood. This understanding is crucial to the determination of the fire origin and the fire cause, i.e. the identification of the 'type of fire'. Fire development, postignition, can be classified in the following way.

After the initial ignition phase, fire enters a second stage known as the incipient phase. The establishment of a self-sustaining chemical reaction characterizes this phase, i.e. unsupported burning occurs. However, although the fire has not yet developed to the point where flaming combustion or even the products of combustion are apparent, there may be sufficient smoke to trigger smoke detectors and there is an increase in the heat generated by the fire. This heat serves to increase the temperature of the burning material, but at this stage does not make a significant contribution to the temperature of the surroundings.

The next stage is emergent smoldering, where the fire grows from a barely self-sustaining reaction to visible flames. Combustion products reach significant levels, and smoke is apparent. The heat loss from the burning material, and the heat generated by the flames causes an increase in the temperature of the surroundings in addition to that of the fuel.

The fourth stage is open burning, where there is a relatively rapid increase in heat output and obvious open flames. As the growing fire heats and ignites nearby material, the room or compartment temperature increases markedly, particularly in the ceiling layer. This layer becomes a source of radiant heat energy, contributing to the rate of temperature rise. There is an adequate supply of air for combustion, and growth of the fire is limited only by the supply of fuel.

If the amount of air in the room or compartment is sufficient, open burning can progress to flashover. For this phenomenon to occur, the heat released, particularly radiant heat from the upper smoke layer, must be sufficient to raise the temperature of all the exposed combustible material in the room to the fire point, at which point all the combustible material in the room will ignite.

The final stage is air- or oxygen-controlled burning. In this stage, the combustion reactions occuring in the room are limited by the lowered proportion of oxygen available. The effect of this restriction is that flames will diminish, possibly to a level of smoldering and, although the temperature may continue to rise, the rate of temperature rise is much reduced. The introduction of additional air/oxygen will return the fire to the open burning situation until the additional oxygen has been consumed.

The increase in burning during this stage will decrease the supply of air in the room. In some situations ventilation will maintain the required airflow, or the fire may breach a window or door and thus

ensure adequate ventilation for continued development. If the proportion of oxygen available to the fire is limited, the fire will revert from open burning to smoldering. This may occur before the fire reaches flashover, in which case flashover may not eventuate.

Although fires generally follow this series of events, the source of ignition can have a significant influence on each of these stages and particularly the time required to complete the early stages.

- A fire that is ignited by self-heating processes or by a relatively low-power electrical fault can take a long time to ignite and may remain in the incipient stage for a considerable time.
- A smoldering fire, such as, but not necessarily confined to, a fire started by self-heating, may move from the emergent smoldering stage to the final air-controlled smoldering stage without entering the open burning phase.
- Progression from emergent smoldering to open burning/flashover to air-controlled smoldering is usually dependent on both the nature of the fuel and the amount of ventilation. Increasing the ventilation initially increases the rate of burning and hence heat output and increases the likelihood of reaching flashover. Beyond a certain upper limit, increasing ventilation has the effect of removing heat from the room and effectively preventing the flashover temperatures being reached.
- A fire involving the ignition of flammable liquid will progress through the incipient stage in an extremely short time, possibly fractions of a second, to reach the open burning stage.

Fire development in the latter circumstances may follow several courses. If the heat generated by the burning liquid is insufficient to ignite the carpet or other combustible materials on, or near where it was sited, the fire might simply cease with the exhaustion of the fuel supply or the air supply. This might result in relatively light burning, isolated areas of scorching or in burn trails in flooring, which is otherwise undamaged.

If ignition of the fuel/air mixture does generate sufficient heat to ignite available combustible materials, this ignition will probably occur at many points on the perimeter of the liquid pool or trail. Any fire that develops from this type of ignition can show comparable levels of damage throughout the premises. Trails related to the spread of flammable liquid are likely to be less distinctive and will certainly become more difficult to distinguish as the fire develops to flashover or oxygen-controlled burning.

Although the fuel vapor/air mixture must be within the explosive limits for ignition to occur, in some instances a substantial proportion of the mixture is within these limits and an ignition known as deflagration can occur. The flame front originating at the point of ignition passes through the mixture at speeds approaching the speed of sound, and significant overpressure may be generated. The subsequent damage can range from minor window glass damage to complete destruction of substantial buildings. Stoichiometric or lean mixtures are more likely to cause major explosions whereas rich mixtures are likely to cause extensive burning with comparatively minor explosion damage.

Although the concepts and principles of fire development remain scientifically valid for all fires, this discussion on fire development applies primarily to fires in compartments. In the case of bushfires (wildfires), concepts such as flash over and ventilation are either irrelevant or insignificant, and the dominating influences on fire development are fuel condition and atmospheric conditions, particularly wind velocity, humidity and ambient temperature.

As in all fire-cause investigations, it is the correct interpretation of the fire travel and development indicators that allows a logical fire cause to be nominated. In the case of bushfires, fuel load, wind speed and wind direction are the primary considerations in the interpretation of fire travel.

Having determined the starting point of the fire, many ignition sources (and therefore, 'types of fire') should be considered. In fact, all the ignition sources mentioned above have been known to be the cause of bushfires. In the case of electrical causes, the formation of sparks caused by clashing electrical conductors (power lines) in high winds is a commonly diagnosed 'type of fire'. Sparks caused by farm machinery, camp fires and acts of arson are all other common ways in which bushfires have started.

By far, however, the most common ignition source for bushfires is lightning strikes. This fire cause is characterized by physical disruption to a tree or other object through which the lightning has 'earthed'. Evidence of a lightning strike is sometimes visible but in other situations, does not survive the ensuing fire.

Whether these types of fire are 'electrical' or 'weather' is academic.

Summary

The basis for the classification of fires outlined above is only one approach. However, it does have the advantage of avoiding any reference to the diagnosis of intent, which with the exception of certain specific cases, is not a scientific conclusion but a forensic outcome. It must be appreciated that the types of

fire discussed are simple examples. In fact, a fire may be a combination of several types (for example, a spark generated by a damaged wheel bearing could ignite a pool of flammable liquid) or, due to environmental factors, might not develop precisely in the manner described (or prescribed). For this reason, the experienced fire investigator will avoid complying with a rigorous classification scheme for fires and will understand that the classification of a fire is achieved after the cause and origin of the fire have been determined, rather than before. A thorough understanding of the various types of ignition and the consequential fire development is crucial to this determination.

See also: **Fire Investigation**: Evidence Recovery at the Fire-scene; Fire-scene; Fire-scene Patterns.

Further Reading

Barnard J and Bradley J (1985) *Flame and Combustion*, 2nd edn. New York: Chapman and Hall.

Bersch W, Holzer G and Sellers CS (1993) *Chemical Analysis for the Arson Investigator and Attorney*, pp. 20–23. Heidelberg: Hüthig Buch Verlag.

Cooke R and Ide R (1985) *Principles of Fire Investigation*. Leicester: Institution of Fire Engineers.

DeHaan J (1991) *Kirk's Fire Investigation*, 3rd edn. Englewood Cliffs, NJ: Prentice-Hall.

International Fire Service Training Association (1985) *Fire Cause Determination*. Oklahoma: Oklahoma State University.

Lecker S (1988) *Incendiaries*. Boulder, Colorado: Paladin Press.

Luke RH and McArthur AG (1978) *Bushfires in Australia*. Canberra: Australian Government Publishing Service.

McCray JA and Cahill TA (1978) *Electronic Circuit Analysis for Scientists*. New York: Wiley.

National Fire Protection Association (1991) *Fire Protection Handbook*, 17th edn. Quincy, MA: National Fire Protection Association.

Redsicker D and O'Connor J (1997) *Practical Fire and Arson Investigation*, 2nd edn. New York: CRC Press.

Selby H (ed.) (1992) *The Aftermath of Death*. Annandale, New South Wales: Foundation Press.

Webster J (1966) *The Complete Australian Bushfire Book*. Melbourne: Nelson.

Yallop HJ (1983) *Fire Investigation*. Solihull, UK: Allen Clift Association.

Yearance R (1987) *Electrical Fire Analysis*, pp. 141–145. Springfield, IL: CC Thomas.

FIREARMS

Contents

Humane Killing Tools
Laboratory Analysis
Range
Residues
Weapons, Ammunitions and Penetration

Humane Killing Tools

S Pollak, University of Freiburg/Br., Freiburg, Germany
P J Saukko, University of Turku, Turku, Finland

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0537

Introduction

A livestock stunner ('humane killer', slaughterer's gun, livestock narcotic device) is used in abattoirs, butcher's shops and farms to induce immediate unconsciousness before slaughtering meat stock. The first stunning devices were constructed at the beginning of the twentieth century. They consisted of a

simple barrel with a heavy flange or trumpet-like muzzle end, which had to be placed against the animal's forehead; an ordinary cartridge carrying a bullet was inserted into the opposite end of the barrel and discharged by striking the firing pin with a wooden hammer. Since the bullets frequently caused accidental injuries, from the 1920s and 1930s onward an increasing number of butchers turned to the safer 'captive-bolt' stunners.

Nevertheless, in several countries such as Britain improved models of free-bullet slaughtering pistols are still used today (for instance the Greener slaughtering guns and the Webley 0.32 in veterinary pistol). The wounds inflicted to the skin by the Greener 'Bell Killer' resemble ordinary gunshot entrance wounds with a variable extent of powder marking, which either reproduces only the central aperture or the outline of the whole muzzle. When using light weapons such as the 'Telescopic' Greener a double pattern of blackening may be seen because of a tendency to bounce at the moment of discharge. The Greener 'Safti-killer' has a patterned muzzle surface in order to prevent slipping. The same is true for the Webley 0.32 in, which otherwise looks like a common pistol; at the distal end of the barrel just short of the muzzle there are two holes as outlets for the combustion gases, which produce corresponding soot deposition in contact shots.

Captive-bolt Humane Killers (Fig. 1)

This type of slaughterer's gun consists of a cylindrical metal tube which serves as a guide for an inlying steel rod ('bolt'), which has a diameter of 7–12 mm. The bolt is driven by the explosion gases of a blank cartridge and propelled a few centimeters beyond the muzzle. The report is less loud than in shots from real guns.

The distal end of the steel rod is usually excavated (conically grooved) so that its front acts like a sharp-edged circular punching tool. The initial velocity of the bolt is rather low ($\sim$40–45 m s^{-1}) compared with hand guns ($\sim$300 m s^{-1}).

After firing, in most models the bolt is automatically drawn back into the barrel either by compressed air or by rubber bushes and/or by a withdrawal spring. This rebound mechanism also prevents the bolt from becoming a free projectile unless the device is manipulated before firing (e.g. by removal of the rubber bush and withdrawal spring). Therefore, injuries can only be inflicted from a short distance of less than the propelled bolt's length ($<$10 cm).

The muzzle plane of the captive-bolt humane killer may show two or four openings arranged symmetrically beside the central hole serving as outlets for the explosion gases which are partly drained off through diverging smoke conduits (**Fig. 2**). This applies to the German Kerner type and some other models (e.g. Dick, Rekord, Ursus, and Kuchen). The number and location of the additional openings essentially determines the smoke soiling pattern in contact or close range shots (paired or clover leaf-shaped soot depositions). In other models (for instance the Cash and the Temple Cox slaughtering guns from the UK or the Schermer type from Germany) the soot emerges only along the bolt's guide so that the blackening from gunpowder residue may either resemble a ring of dirt encircling the entrance hole or cover a greater area in circumferential distribution. The extent of smoke deposition depends on the propellant (black or smokeless powder) and on the strength of the charge used.

The blank cartridges (**Fig. 3**) are loaded with smokeless or black powder and differ with regard to the

Figure 1 Two makes of captive-bolt stunners. Top: Schermer type (mod. ME, cal. 10 × 11 mm, diameter of the bolt 12 mm); bottom: Kerner type (mod. 289, cal. 9 × 17 mm, diameter of the bolt 11 mm).

Figure 2 Muzzle end of two captive-bolt stunners. Left: Kerner type with two opposite openings (outlets for the explosion gases; *arrows*); right: Schermer type without additional smoke conduits.

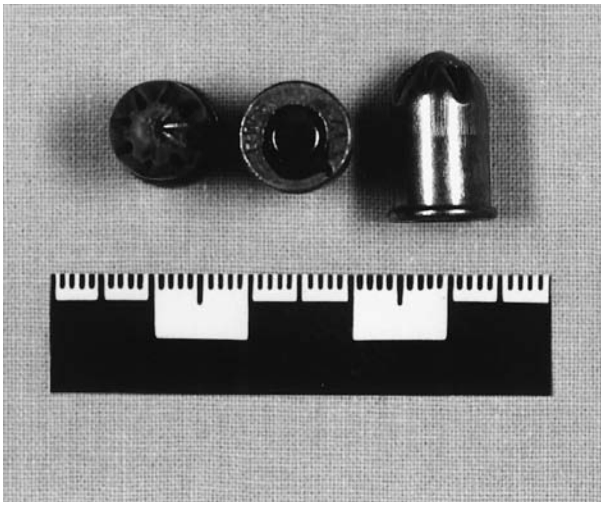


Figure 3 Blank ammunition for captive-bolt stunners. Cattle-killing cartridges RWS 9 × 17 mm; the case mouth is sealed by crimping the metal and inserting a wax plug. The strength of the cartridge is designated by a green color marking on the base of the case ('green' meant for small cattle and pigs).

caliber fitting into the chamber of the respective model (e.g. 0.22 in, 9 × 17 mm, 10 × 11 mm). The strength of ammunition varies depending on the animal to be slaughtered. The manufacturers offer several sizes of charge distinguished by color markings on the base of the cartridge case (with the strongest ones meant for heavy bulls painted black or red).

Injuries from Captive-bolt Narcotic Devices

When stunning meat stock the muzzle has to be placed tightly against the forehead so that the whole length of the bolt enters the cranial cavity. Analogous to the veterinary application, in most human victims the weapon is fired in contact with the skin. Nevertheless, the entrance wounds lack all the classical signs of contact shots as known from conventional firearms. There is no muzzle imprint and no stellate splitting because the skin is not ballooned out by expanding combustion gases, which follow a free bullet but not a captive bolt. Consequently, only little, if any, soot is found in the depth of the entrance hole. The excavated distal end of the bolt causes a sharp-edged circular punch lesion with the diameter of the entrance hole being a little smaller than that of the bolt due to the skin's elasticity. If the shot is fired at an oblique angle there is a unilateral sickle-shaped beveling of the corium indicating the bolt's direction.

Humane killers with two or four smoke outlets on the muzzle end produce a characteristic soot pattern consisting of roundish or elliptic zones of blackening arranged in congruity with the respective openings. Variations in shape and location of these foulings point to an oblique holding of the instrument. In noncontact shots the smoke deposits become larger, but less intensive, with increasing distance and finally disappear. Interposed hair or garments covering the entry site may prevent the skin being blackened at all. As mentioned above, other kinds of captive-bolt instruments do not have any additional smoke conduits; in such cases the entrance wound is surrounded by a ring or a circumferential area of powder soiling. If the soot was wiped off in the course of surgical treatment, the affected skin areas tend to dry up producing a brownish discoloration.

The typical entry sites are regions of the body where only a thin layer of soft tissue overlies flat bones (forehead, temple, parietal and occipital region). When the bolt enters the skull, it produces a round hole, which is sharp-edged ('clean-cut') on the outer table and beveled-out ('cratered') on the inner table. The external aspect of the entrance hole almost exactly reflects the cross-section of the bolt, whereas the inner diameter of the bone defect is much larger. In shots fired at an acute angle the entrance in bone is oval-shaped and sometimes accompanied by semicircular frame-like fracture lines.

Due to the limited length of the bolt, in shots to the head no exit wound is to be expected. Of course there is no bullet left at the end of the wound track. Because of the bolt's punching function the impact surface is forced into the depth of the brain. The punched-out skin and bone are thrust inwards and left there as a so-called 'imprimatum'. The skin is nearly always a component of the imprimatum (**Fig. 4**) and must be regarded as bacterially contaminated and therefore infectious (Fig. 4). Apart from the skin and the bone imprimatum, other material from outside such as pieces of hairs and textiles may be pushed into the wound path. Despite its low velocity the bolt causes a remarkably wide wound channel owing to the large cross-section and the excavated front; hemorrhage results from mechanical tissue damage with disruption of vessels. In contrast to the extensive local traumatization there are no morphological signs of elevated intracranial pressure from temporary cavitation such as indirect fractures, cortical contusion and intracerebral extravasation away from the permanent tract.

Fatalities from Humane Killers

Most of the fatal injuries from livestock narcotic devices are suicides, which account for 85% of the published statistical figures. Though the number of accidents and homicides is comparatively small, the

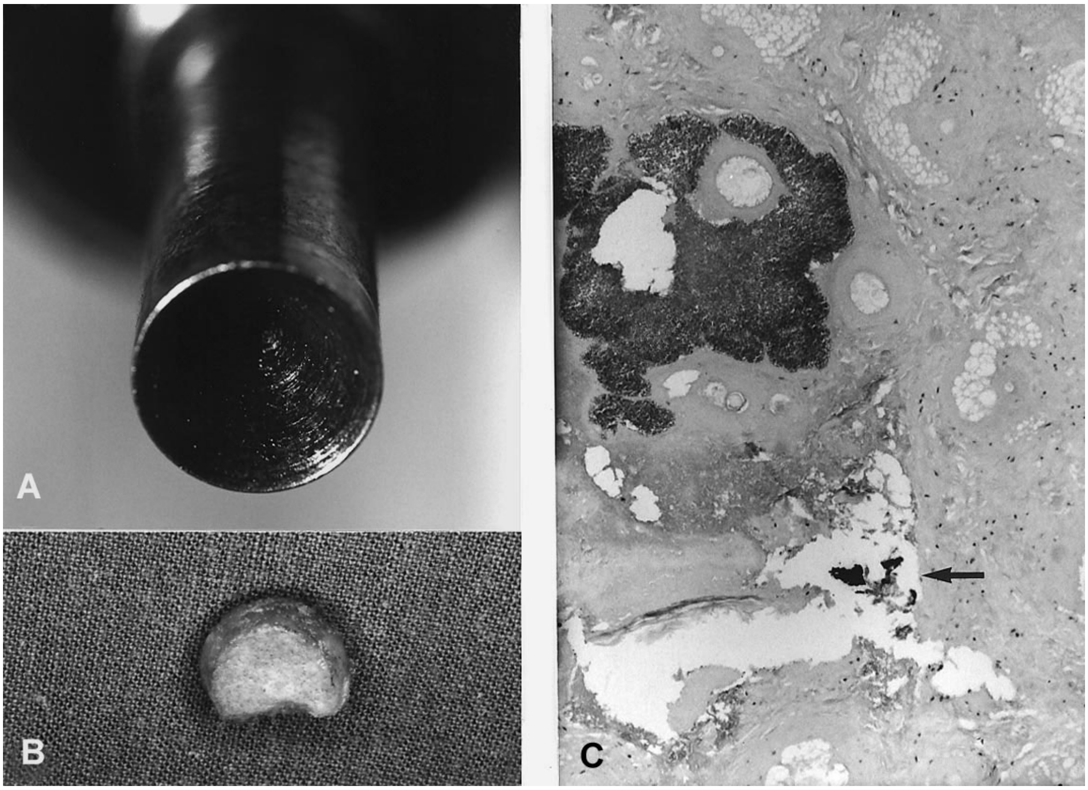


Figure 4 (**A**) Distal end of the captive-bolt in a Kerner type slaughtering gun; the steel bolt is conically grooved. (**B**) Punched-out piece of skin from the depth of the intracranial wound path ('skin imprimatum'). (**C**) Histological section of a skin imprimatum showing a bacterial colony and black powder residues (*arrow*). Hematoxylin & eosin stain.

possibility of such incidents has to be kept in mind. In the German language literature there are reports on more than a dozen homicide cases with a high percentage of female victims. In contrast to this, the great majority of suicides are males, most of them butchers or farmers, familiar with slaughtering guns. The frequency curve attains its peak in the fifth and sixth decade of life.

The most common sites of suicidal entrance wounds are, in decreasing order of occurrence: the forehead (**Figs 5** and **6**), the temple, the parietal and occipital region, the precordium and the mouth. Some individuals use two or more different methods when committing suicide, for example by placing a noose around the neck before shooting themselves in the head with a livestock stunner. Some authors have even described suicide by repeated cranial shots from captive-bolt humane killers. Such observations prove that bolt injuries of the brain are not necessarily followed by immediate or persistent unconsciousness, if the bolt does not strike a vital area of the brain. Despite severe damage to the frontal and/or temporal lobes the individual may be capable of reloading the humane killer and firing a second time.

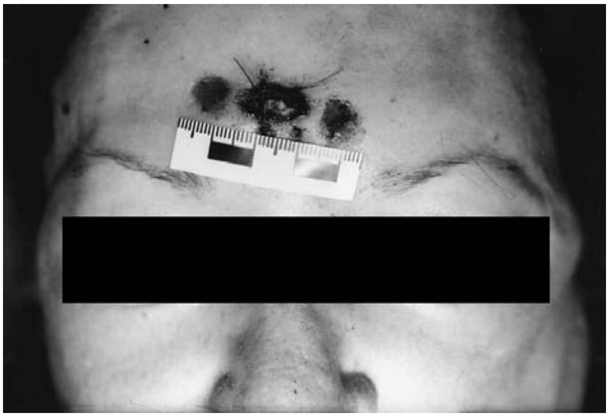


Figure 5 Contact entrance wound from a Kerner type captive-bolt livestock stunner in the central forehead of a suicide. The initially circular skin hole is partly held together by a surgical stitch. The entrance wound is accompanied by two roundish soot deposits corresponding to the openings of the smoke conduits.

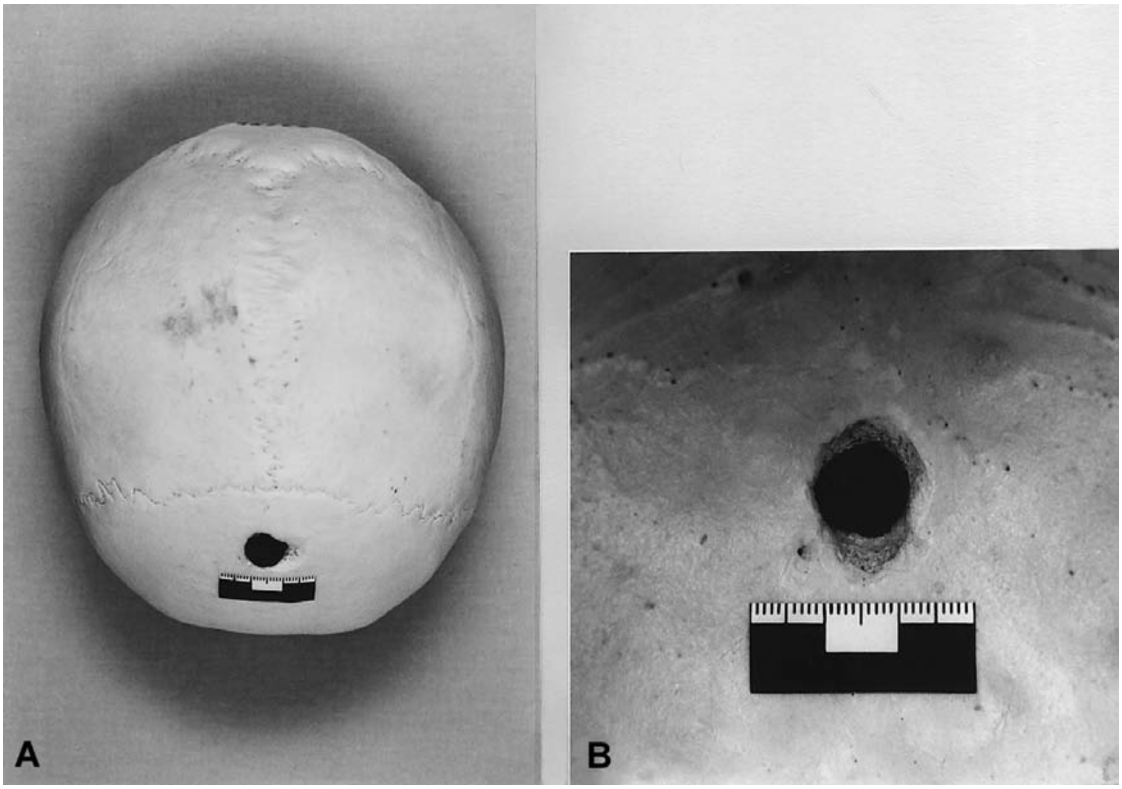


Figure 6 (**A**) Skull-cap of a suicide who shot himself in the forehead with a captive-bolt stunner. The punched-out entrance hole is located in the frontal squama. (**B**) Internal aspect of the skull showing a cone-shaped crater that is appreciably larger than the diameter of the bolt.

In fatal cases the survival period varies widely and may last up to several weeks or even months. The mortality has decreased, thanks to the great progress made in neurosurgery and intensive care, but in general the prognosis is still rather poor, though a recent clinical study reports a survival rate of more than 50%. In addition to extensive brain damage and intracranial bleeding from lacerated vessels, inflammatory complications such as pyogenic encephalomeningitis or a cerebral abscess may be the cause of a delayed fatal outcome.

Suicidal injuries from captive-bolt humane killers are almost always contact shots. In most cases the shooting device is found close to the deceased. The hand holding the muzzle end often reveals soot on the radial surface of the index finger and on the ulnar surface of the thumb. Other visible soot deposits may be transferred to the hands in the course of loading. In suicides firing again after the first shot the hands and the weapon are usually soiled with blood pouring from the entrance wound in the course of reloading.

The relatively small number of homicide cases is probably due to the limited length of the bolt which does not allow shots from a distance of more than a few centimeters. Most victims are, therefore, either defenseless or unaware of an attack.

Stud Guns

In this context powder-actuated fastening tools are only mentioned insofar as it serves to distinguish between stud guns and humane killers. Stud guns are similar to ordinary firearms and use blank cartridges to fire pins, threaded studs, nails or fasteners into wood, concrete, masonry or metal structures, so that other materials or objects may be affixed to the receiving part. The blank cartridges range in caliber from 0.22 in to 0.38 in and are loaded with fast-burning propellants. The external appearance of a stud gun resembles a large pistol with a flat face plate attached to the muzzle, which has to be pressed firmly against a solid surface before it can be fired. The purpose of this built-in safety mechanism is to prevent free flight or ricochet of the stud. Nevertheless, numerous accidents and a few suicides have occurred with stud guns.

A major cause of injuries and deaths is perforation of flimsy walls with the result that a worker behind or a bystander is struck by the nail. Many accidents are due to ricochet of the steel missile which as a result becomes bent. Thus the finding of a bent nail on radiographs of the victim can be helpful in determining the circumstances of the accident. In suicides, the most common sites for a stud gun entrance wound are the chest, the forehead, and the temple.

The morphology of skin wounds from studs is variable, e.g. a distance shot with a straight and pointed steel missile causes an inconspicuous slit like puncture wound, whereas a bent stud after ricochet produces an irregular, jagged laceration with abraded margins. In (suicidal) contact shots stellate entrance wounds surrounded by muzzle imprint marks have been observed.

See also: **Firearms**: Weapons, Ammunitions and Penetration; Range; Residues.

Further Reading

Betz P, Pankratz H, Penning R and Eisenmenger W (1993) Homicide with a captive bolt pistol. *American Journal of Forensic Medicine and Pathology* **14**:54–57.

DiMaio VJM (1985) *Gunshot Wounds*, pp. 233–234. New York, Amsterdam, Oxford: Elsevier.

Goldin MD and Economou SG (1965) Stud gun injuries. *Journal of Trauma* **5**:670–677.

Hunt AC and Kon VM (1962) The patterns of injury from humane killers. *Medicine, Science and the Law* **2**:197–203.

Janssen W and Stieger W (1964) Verletzungen durch Bolzenschuss-Apparate unter besonderer Berücksichtigung der Spurenmerkmale. *Archiv für Kriminologie* **134**:26–37, 96–102. (in German).

Knight B (1996) *Forensic Pathology*, 2nd edn, p. 269. London: Arnold.

Lignitz E, Koops E and Püschel K (1988) Tod durch Bolzenschussgeräte – Eine retrospektive Analyse von 34 Fällen aus Berlin und Hamburg. *Archiv für Kriminologie* **182**:83–93. (in German).

Opeskin K and Cordner S (1990) Nail-gun suicide. *American Journal of Forensic Medicine and Pathology* **11**:282–284.

Pollak S (1977) Morphology of injuries by 'humane killer' (livestock stunner). *Zeitschrift für Rechtsmedizin* **80**:153–165. (in German)

Pollak S and Maurer H (1987) On the clinical importance of the imprimatum in injuries from slaughterer's gun (livestock narcotic device, 'humane killer'). *Acta Chirurgica Austriaca* **19**:29–37. (in German).

Pollak S and Reiter C (1981) 'Bolt projectiles' discharged from modified humane killers. *Zeitschrift für Rechtsmedizin* **87**:279–285. (in German)

Polson CJ (1965) *The Essentials of Forensic Medicine*, 2nd edn, pp. 183–185. Oxford: Pergamon Press.

Sellier K (1982) *Schusswaffen und Schusswirkungen I*, 2nd edn, pp. 120–126, 292–293, 393, 488–491. Lübeck: Schmidt-Römhild. (in German).

Spitz WU (1993) Stud guns (nail guns). In: Spitz WU (ed.) *Medicolegal Investigation of Death*, 3rd edn, p. 373. Springfield: Charles C Thomas.

Weedn VW and Mittleman RE (1984) Stud guns revisited: report of a suicide and literature review. *Journal of Forensic Science* **29**:670–678.

Laboratory Analysis

W F Rowe, The George Washington University, Washington, DC, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0538

Introduction

Firearms examiners conduct a variety of laboratory examinations. They identify and compare fired bullets and cartridges; they examine powder and shotgun pellet patterns; they restore obliterated serial numbers and they conduct toolmark examinations.

Examination of Weapons

For chain-of-custody purposes the complete identification of a firearm requires the following information: make or manufacturer, type (revolver, semiautomatic pistol and so forth), caliber, serial number, the model, the number of shots and the barrel length. Much of this information is stamped on the frame of the weapon. Many internal components of a firearm may also bear partial serial numbers; these should also be checked to determine if any components have been replaced. Information stamped on a firearm should not be taken at face-value. A weapon may have been altered; it may also be a 'bootleg' weapon, i.e. a copy of a commercially produced weapon made by an unknown gunsmith. Some bootleg weapons even have duplicated inspectors' proofmarks.

When weapons are submitted for microscopic comparison of their test-fired bullets with questioned bullets a number of preliminary examinations should be done. The exact nature of these will depend on the nature of the case being investigated; however, the preliminary examinations could include processing for fingerprints, examination of the interior of the muzzle of the weapon for blood and other tissues, determination of trigger pull and examination of the

functioning of the weapon's firing mechanism and safeties.

Processing for fingerprints will usually be the first of these examinations carried out. Firearms frequently do not yield fingerprints. A light coating of gun oil will obscure latent fingerprints and surfaces such as knurled grips are not conducive to the deposition of useable latent fingerprints. Nevertheless, the value of fingerprints for identification purposes is so great that latent fingerprint development should be routinely carried out when a weapon is not found in the possession of a suspect.

The bore of the weapon can be inspected with a helixometer. Contact shots, particularly to the head, can result in the blowback of blood and tissue into the muzzle. Other trace evidence may also be observed. If a shooter carries his weapon in a pocket, fibers from the pocket may find their way into the gun bore. The examination of the gun bore with a helixometer may be valuable from another perspective. It may reveal significant corrosion which may preclude meaningful microscopic comparison of test-fired bullets.

The determination of the weapon's trigger pull is carried out with a spring gauge or by suspending increasing weight on the trigger until the firing mechanism is actuated. Some revolvers can be fired either single action (the firing mechanism is thumb-cocked and the weapon is then fired by squeezing the trigger) or double-action (a single long trigger pull rotates the cylinder, raises the hammer and then actuates the firing mechanism). Double-action trigger pulls are much larger than single-action trigger pulls; if a weapon can be fired either way both trigger pulls must be determined. The determination of trigger pulls is both a safety measure and an investigative procedure. The firearms examiner will generally test-fire the weapon and needs to be aware if the weapon's low trigger pull makes it prone to accidental discharge. The suspect in a shooting investigation may allege that the shooting was due to an accidental discharge resulting from a low trigger pull. In cases of suspected suicide the trigger pull is also an issue: could the decedent have shot himself accidently while cleaning the weapon or did firing the weapon require a deliberate act of will?

The examination of the functioning of a weapon's firing mechanism and safety features is both a safety measure for the firearms examiner and an investigative procedure. The firearms examiner needs to know if a weapon can be safely handled and fired before an attempt is made to test fire it. If the weapon as submitted by police investigators cannot be fired (due to missing or damaged components) that fact must be noted and the weapon rendered operable by replacement of components from the firearms laboratory's firearms reference collection. Many weapons lack safety features, whereas others have several. The Colt M1911A1 semiautomatic pistol is an example of a weapon with several safety features. It has a grip safety which must be depressed by squeezing the grip with the firing hand before the weapon will fire. A manual safety catch on the left rear of the frame in the up or safe position both blocks the trigger and locks the pistol's slide in the forward position. Finally, the hammer has a half-cock position. If when the pistol is thumbcocked the hammer slips before the full-cock position is reached the hammer will fall only to the half-cock position and not discharge the pistol.

It may be necessary to verify the caliber of the weapon. This can be done with precision taper gauges or by making a cast of the interior of the barrel. Other general rifling characteristics (e.g. number of lands and grooves, direction of twist of the rifling) can also be determined from a cast but are more conveniently determined from test-fired bullets.

Examination of Questioned Bullets and Cartridges

In the course of a preliminary examination of questioned bullets and cartridges the firearms examiner should note all identifying marks placed on these items by police investigators. He should also note the presence of any trace evidence on the questioned bullets. Bullets may have blood or other tissue on them. Bullets may also pick up fibers when passing through wood, clothing or other textiles. If a bullet has ricocheted off a surface there may be fragments of concrete or brick embedded in it. These bits of trace evidence may be very helpful in the later reconstruction of the shooting incident under investigation. Any patterned markings on the bullet should also be noted. These may also be significant in reconstruction of a shooting incident. Bullets have been found bearing impressions of the weave of clothing, wires in a window screen and police badges. Cartridge cases can be processed for latent fingerprints. Bluing solutions have proven useful for developing latent fingerprints on cartridge cases.

When fired bullets or cartridges are submitted for examination the firearms examiner must make an initial determination of the class characteristics of the weapon that fired them. From these class characteristics the possible makes and models of firearm that fired the bullet or cartridge can be determined.

Determination of general rifling characteristics from fired bullets

Caliber The caliber of an undamaged bullet can be determined in one of two ways: the diameter of the bullet may be measured with a micrometer; or the

questioned bullet can be compared with bullets fired from weapons whose general rifling characteristics are known. If the bullet is severely deformed but unfragmented it can be weighed. This may not specify the particular caliber but will eliminate certain calibers from consideration. The weight of the bullet is also a class characteristic of the brand of ammunition. If the bullet is fragmented its possible caliber may still be determined if a land impression and an adjacent groove impression can be found. The sums of the widths of a land and impression and an adjacent groove impression have been tabulated for a large number of different makes and models of firearms. The specific caliber of the bullet will not in general be determined, but many calibers will be eliminated.

Number of lands and grooves Unless a fired bullet is fragmented the number of lands and grooves can be determined by simple inspection.

Widths of the lands and grooves These may be determined using a machinist's microscope or by comparison of the bullet with bullets fired from weapons having known general rifling characteristics.

Direction of twist of the rifling If the fired bullet is relatively undistorted the direction of twist of the rifling may be determined by noting the inclination of the land impressions with respect to the axis of the bullet: land impressions inclined to the right indicate right twist rifling, whereas land impressions inclined to the left indicate left twist rifling.

Degree of twist of the rifling This can be determined using a machinist's microscope or by comparison of the bullet with bullets fired from weapons having known general rifling characteristics.

Determination of class characteristics from cartridges

The following class characteristics may be determined from fired cartridges: caliber; type of cartridge; shape of firing chamber; location, size and shape of the firing pin; sizes and locations of extractors and ejectors. These may be determined by visual inspection with the naked eye or with a low-power microscope.

Pitfalls in determining class characteristics from fired bullets and cartridges

Care should be taken in inferring the class characteristics of a firearm from fired bullets and cartridges. A number of devices have been developed to permit the firing of ammunition in weapons for which the ammunition was not designed. These vary from the very simple to the very complex. An example of the very simple is the crescent-shaped metal tabs inserted in the extractor grooves of 0.45 ACP cartridges (for which the Colt 0.45 caliber M1911A1 semiautomatic pistol is chambered) to enable the cartridges to be loaded in 0.45 caliber revolvers. An example of a more complex device would be a barrel insert that allows 0.32 caliber pistol cartridges to be chambered and fired in a shotgun. Field expedients can also be used to fire subcaliber rounds from a weapon. Subcaliber cartridges can be wrapped in paper and then loaded into the chambers of a revolver. Finally, some weapons can chamber and fire cartridges of the wrong caliber. The 7.62 mm Luger can chamber and fire 9 mm parabellum rounds.

Observations of individual characteristics

In order to determine whether a questioned bullet was fired from a particular firearm the questioned bullet must be compared with bullets that have been fired from the weapon that is suspected of firing the questioned bullet. In the very early days of firearms examinations (prior to the mid-1920s) test bullets were obtained by pushing a bullet through the barrel of the weapon. However, this procedure ignores the fact that when a bullet is fired down the barrel of a firearm it may undergo significant expansion (the technical term is obturation); as a result fired bullets may be marked differently by the rifling than bullets merely pushed down the barrel. Test-fired bullets are now shot into some type of trap that permits the bullets to be recovered in a minimally distorted condition. Some laboratories use horizontal or vertical water traps. Metal boxes filled with cotton waste have also been used as bullet traps; however, cotton fibers can be abrasive to the surfaces of hot lead bullets.

The comparison of bullets and cartridges is carried out with a comparison microscope. This actually consists of two compound microscopes connected by an optical bridge so that the images produced by each microscope can be viewed side by side. In the case of fired bullets, the firearms examiner attempts to find matching patterns of striations on both the questioned and test-fired bullets (**Fig. 1**). The probability of an adventitious match declines as increasing numbers of striations are matched. The probability of a random match of a pattern containing five or more closely spaced striations is negligible; however, further research is required to establish objective standards for declaring a match.

The firearms examiner can look at a variety of marks in making his comparisons of bullets: land impressions, groove impressions (often absent or insignificant on jacketed bullets), skid marks (marks made by the leading edge of the rifling as the bullet

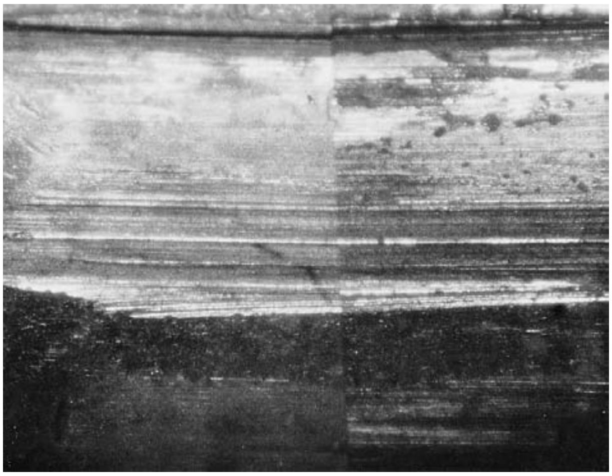


Figure 1 Matching striation patterns in the land impressions of two bullets fired in the same weapon.

enters the rifling), shaving marks (marks made on revolver bullets when the bullets strike the edge of the forcing cone, the flared opening at the breech end of a revolver barrel) and slippage marks (marks produced when an overlubricated bullet slides down the barrel without engaging the rifling). Matching patterns of marks are also sought by the firearms examiner by comparing firing pin impressions, breechblock markings, extractor marks, ejector marks and magazine marks (**Fig. 2**). Extractor marks, ejector marks and magazine marks are striated marks like land and groove marks on bullets. Firing pin impressions and breechblock marks may also contain striations if these components of the firearm were finished by filing.

Firearms examiners' conclusions

Firearms examiners can render one of three opinions following the microscopic comparison of fired bullets and/or cartridges.

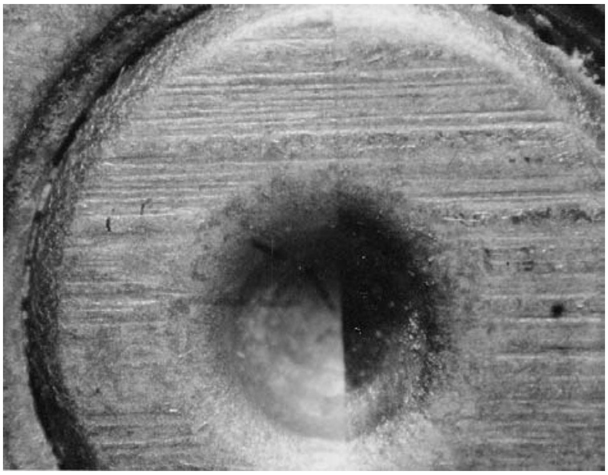


Figure 2 Breech block impressions on the primers of two centerfire cartridges fired in the same pistol.

1. Positive identification. The questioned firearm fired the questioned bullets and/or cartridges. In this case both the class and individual characteristics of the questioned bullets and/or cartridges match those of test-fired bullets and/or cartridges from the questioned firearm.
2. Negative identification. The questioned firearm did not fire the questioned bullets and/or cartridges. The class characteristics of the questioned firearm do not match the class characteristics of the firearm that fired the questioned bullet and/or cartridges.
3. Inconclusive. The questioned firearm could have fired the questioned bullets and/or cartridges. The class characteristics of the questioned firearm match those of the firearm that fired the questioned bullets and/or cartridges but the individual characteristics do not match. The failure of the individual characteristics to match can result from a number of circumstances. First of all, the questioned bullets and/or cartridges could have been fired in a different firearm having the same class characteristics as the questioned firearm. On the other hand the barrel of the questioned firearm may have undergone substantial alteration due to corrosion (e.g. rusting) or use (e.g. firing of a large number of rounds). Criminals may also attempt to forestall firearms examinations by replacing firing pins, filing breechblocks or altering the interior of the barrel with abrasives.

Pitfalls in comparing bullets and cartridges

Experience with the comparison of bullets fired from sequentially rifled barrels has revealed that the land and groove markings on fired bullets consist of (a) accidental characteristics (which vary from weapon to weapon and are true individual characteristics), (b) subclass characteristics (usually reflecting the particular rifling tool used to rifle the barrels) and (c) class characteristics. The problem for the firearms examiner is distinguishing type (a) marks from type (b) marks. The interpretive problem is made worse by the fact that type (b) marks are often more prominent than type (a) marks. This circumstance can lead the inexperienced firearms examiner to erroneously conclude that questioned bullets were fired from a particular weapon when in fact they were fired from a weapon rifled with the same rifling broach.

With new weapons firearms examiners have observed a 'setting-in' period during which sequentially fired bullets may not match one another. There is a period of use during which individual characteristics produced by the rifling operation change or even disappear. Eventually, the pattern of individual

characteristics settles down and sequentially fired bullets may be readily matched. If the firearms examiner suspects that the individual characteristics in the land and groove marks produced by a weapon may not be reproducible from round to round either because the weapon is new or because the gun bore is badly corroded, a number of bullets can be test-fired through the weapon to try to match their individual characteristics before comparing them with the questioned bullets. If the test-fired bullets cannot be matched to each other there is little point in attempting to match them to the questioned bullets.

Firing pin impressions and boltface markings can also present similar interpretive patterns. Coarse markings resulting from the manufacturing process may be the same on different firing pins or breechblocks; the true individual characteristics may be very fine markings that are easily overlooked. Some weapons (e.g. 0.25 caliber auto pistols) produce very poor breechblock markings in which only one or two true individual characteristic marks may be present. Breechblock markings may be of such poor quality that no conclusions can be drawn from their comparison; in such an instance the firearms examiner may have to make use of extractor or ejector markings.

Recent Advances in Firearms Comparisons

The advent of modern high-speed digital computers has caused major changes in the field of firearms examination. At first, existing information on rifling characteristics and other class characteristics of firearms was incorporated into computerized databases that could be quickly searched for makes and models of firearms that could have fired a questioned bullet or cartridge. More recently two computer systems have been developed to capture, store and compare images of fired bullets and cartridges. DRUGFIRE® is a multimedia database imaging system developed by the Federal Bureau of Investigation and Mnemonics Systems Inc. This system allows the firearms examiner to capture magnified images of fired bullets and cartridges; the computer system can then compare these images with ones already stored in the system to find possible matches. The firearms examiner can then conduct a side-by-side comparison of the images to determine if indeed the items of evidence do match. IBIS (Integrated Ballistics Identification System) was developed by the Bureau of Alcohol, Tobacco and Firearms (BATF) of the US Treasury Department and Forensic Technology of Canada. IBIS consists of two modules: BULLETPROOF™ for the comparison of the images of fired bullets and BRASSCATCHER™ for the comparison of fired cartridges. Like DRUGFIRE®, IBIS can search for matches among previously stored images and allows the firearms examiner to compare the images of possibly matching bullets and/or cartridges. The formats of the DRUGFIRE® and IBIS systems are not compatible with one another so that images captured by one system cannot be directly imported into the other system. The great advantage of the two systems compared to conventional firearms examinations is that they allow the firearms examiner to rapidly compare bullets and cartridges test-fired from weapons seized by police with evidentiary bullets and cartridges from open shooting cases.

Miscellaneous Examinations

Firearms (as well as many other manufactured articles) bear stamped serial numbers. The complete serial number is stamped on the frame of the weapon, usually directly above the trigger; partial serial numbers may also be stamped on other components of weapons (e.g. barrels). Criminals are aware that serial numbers can be used to trace a firearm to its last legal owner, and consequently, may attempt to obliterate the serial number by filing or grinding. If the filing or grinding does not remove too much metal, enough of the disordered metallic crystal structure produced by stamping the serial number may remain to permit partial or complete restoration of the serial number. Stamped serial numbers may be restored by ultrasonication, chemical or electrochemical etching or by the Magnaflux™ method. Ultrasonication uses liquid cavitation to eat away the metal surface; the disordered crystal structure that underlies the stamped serial number is eaten away faster. Chemical and electrochemical etching also rely on the more rapid dissolution of the disordered metal structure. The Magnaflux™ method depends on the concentration of magnetic lines of force by the disordered crystal structure. Magnaflux™ powders consist of magnetic particles which cluster where the magnetic lines of force are more concentrated. Unlike ultrasonication and chemical and electrochemical etching, the Magnaflux™ method is nondestructive; however, its use is restricted to restoration of stamped serial numbers on ferromagnetic alloys.

See also: **Firearms**: Weapons, Ammunitions and Penetration; Range; Residues. **Pattern Evidence**: Shotgun Ammunition on a Target.

Further Reading

Berg SO (1977) The forensic ballistics laboratory. In: Tedeschi CG, Eckert WG and Tedeschi LG (eds)

Forensic Medicine: A Study in Trauma and Environmental Hazard, vol. 1. Philadelphia: WB Saunders.

Biasotti AA (1959) A statistical study of the individual characteristics of fired bullets. *Journal of Forensic Science* **4**:35–50.

Booker JL (1980) Examination of the badly damaged bullet. *Journal of the Forensic Science Society* **20**:153–162.

Ceccaldi PF (1962) Examination of firearms and ammunition. In: Lundquist F (ed.) *Methods of Forensic Science*, vol. 1. New York: Interscience Publishers.

Davis JE (1958) *An Introduction to Tool Marks, Firearms and the Striagraph*. Springfield: Charles C Thomas.

Dougherty PM (1969) Report on two early United States firearms identification cases. *Journal of Forensic Sciences* **14**:453–459.

Hamby JE and Thorpe JW (1997) Firearms reference collections – their size, composition and use. *Journal of Forensic Sciences* **42**:461–465.

Hatcher JS, Jury FJ and Weller J (1957) *Firearms Investigation, Identification and Evidence*. Harrisburg: Stackpole Books.

Krcma V (1961) The identification of pistols by serial numbers and other markings. *Journal of Forensic Sciences* **6**:479–500.

Krcma V (1996) Fluted and annular grooved barrel chambers in firearms. *Journal of Forensic Sciences* **41**:407–417.

Mathews JH (1962) *Firearms Identification*. three volumes. Springfield: Charles C Thomas.

Munhall BD (1960) Firearms identification problems pertaining to supplemental chambers, auxiliary cartridges, insert barrels and conversion units. *Journal of Forensic Sciences* **5**:319–330.

Nichols RG (1997) Firearm and toolmark identification criteria: a review of the literature. *Journal of Forensic Sciences* **42**:466–474.

Rowe WF (1988) Firearms identification. In: Saferstein R (ed) *Forensic Science Handbook*, vol. II. Englewood Cliffs: Prentice Hall.

Rowe WF (1991) Statistics in forensic ballistics. In: Aitken CGG and Stoney DA (eds) *The Use of Statistics in Forensic Science*. Chichester: Ellis Horwood.

Sojat SG (1965) Organization, study and use of fired standards. *Journal of Forensic Science* **10**:1–22.

Thomas F (1967) Comments on the discovery of striation matching and on early contributions to forensic firearms identification. *Journal of Forensic Sciences* **12**:1–7.

Turley DM (1987) Restoration of stamp marks on steel components by etching and magnetic techniques. *Journal of Forensic Sciences* **32**:640–649.

Uchiyama T (1975) A criterion for land mark identification. *Reports of National Research Institute of Police Science* **28**:179–187.

Uchiyama T (1988) A criterion for land mark identification. *AFTE Journal* **20**:236–251.

Uchiyama T (1992) The probability of corresponding striae in toolmarks. *AFTE Journal* **24**:273–290.

Wolfer DA and Lee WJ (1960) Application of magnetic principles to the restoration of serial numbers. *Journal of Criminal Law, Criminology and Police Science* **50**:519.

Young SG (1974) The restoration of obliterated stamped serial numbers by ultrasonically induced cavitation in water. *Journal of Forensic Sciences* **19**:820–835.

Range

W F Rowe, The George Washington University, Washington DC, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0534

Introduction

Firearms examiners are frequently called on to estimate the range from which a gunshot was fired. The distribution of gunshot residue on the target surface can be used for this determination. Gunshot residue consists of the following materials: (1) unburned propellant particles; (2) partially burned propellant particles; (3) soot from the combustion of propellant; (4) nitrates and nitrites from combustion of propellant; (5) particles of primer residue (oxides of lead, antimony and barium); and (6) particles of bullet or bullet jacket. The following terminology is frequently used to characterize ranges of fire based on the deposition of gunshot residue:

- Distant shot: the shot was fired from such a distance that no gunshot residue reached the target surface. Distant gunshot wounds are circular or elliptical defects surrounded by a marginal abrasion or contusion ring caused by the bullet stretching and tearing the skin. The contusion ring is frequently obscured by a gray ring of bullet ‘wipe’, consisting of lubricant and gunpowder combustion products. Distant bullet holes in clothing, walls and the like will lack the contusion ring but will usually show a ring of bullet ‘wipe’. An elemental analysis of the residue surrounding the hole may be required to determine whether a hole in an item of clothing, a door or a wall is in fact a bullet hole.
- Close-range shot: the shot was fired close enough to the target surface for some gunshot residue to reach it. For handguns close range typically means within 12–18 inches (30–45 cm). However, this is very much dependent on the type of ammunition being fired and the condition of the weapon. For rifles and shotguns close range typically means within several feet. The gunshot residue deposited on the target

surface consists primarily of unburned propellant particles, partially burned propellant particles and soot. The unburned and partially burned propellant particles produce what is called 'stippling' or 'tattooing'. The heated particles may be fused to the surface of the target. The particles may be driven into exposed skin (hence the name 'tattooing'). As the distance between the gun muzzle and the target surface decreases the gunshot residue pattern becomes smaller and more concentrated, with increasing amounts of stippling and soot deposition. Gunshot residue patterns display a variety of patterns, such as irregular, circular or petal.

- Near-contact shot: the shot was fired at a range of 1 in (2.5 cm) or less. At this range, there will be a concentrated gunshot residue pattern. The muzzle flash (the incandescent gases issuing from the muzzle) will also interact with the target surface. Hair will be singed. Textile fibers may be singed or even melted and the textile structure will be disrupted. Woven textiles will split apart in an X-shaped pattern; knitted textiles will show a large circular or elliptical hole.
- Contact shot: the shot was fired with the muzzle in contact with the target surface. It is useful to distinguish loose contact shots from hard contact shots. A loose contact shot results when the muzzle of the firearm just touches the target surface. Gunshot residue may be blown out along the target surface or between layers of clothing. A hard contact shot results when the muzzle is pressed tightly against the target surface. Gunshot residue tends to follow the bullet into the target and not soil the target's surface. The heat of the muzzle flash can scorch or melt textile fibers, producing a so-called ironing effect. Hard-contact gunshot wounds over bony plates (e.g. the vault of the skull) can produce characteristic stellate defects. These are produced when the hot propellant gases create a pocket between the overlying soft tissue and the bone (depositing gunshot residue on the surface of the bone). The soft tissue is pressed against the weapon's muzzle, producing a patterned contusion. If the gas pressure is high enough the soft tissue can split open; blood and tissue may be blown back into the weapon's muzzle. The soft tissue in loose and hard contact gunshot wounds frequently displays the cherry red color of carboxyhemoglobin, which is produced by the reaction of hemoglobin with carbon monoxide in the propellant gases.

In order to conduct a range-of-fire determination the firearms examiner requires certain pieces of evidence or information. He should have the original gunshot residue pattern, the weapon believed to have fired the pattern, ammunition from the same lot as that used to fire the original gunshot residue pattern and knowledge of the weather conditions prevailing at the time of the shooting. In the case of a gunshot residue pattern on skin a scaled black and white or color photograph of the pattern is usually used for comparison purposes. However, black and white photographs may be misleading as some of the discoloration of the skin in a powder pattern is the result of a vital reaction of the skin to the impact of hot propellant grains. Even a color photograph may be misleading. Skin blemishes may be mistaken for stippling or tattooing. Ideally, the firearms examiner should be present at the autopsy of a shooting victim so that he can examine the gunshot residue pattern himself. Older homicide investigation texts frequently recommended excising the skin bearing gunshot residue patterns at autopsy and sewing the formalin-fixed tissue to a metal hoop. The advisability of this practice is open to question: the formalin-fixed tissue will shrink to some degree; the evidence may be deemed by a judge too inflammatory to be admitted at trial; and members of some religions would regard the excision of the powder pattern as an unconscionable mutilation of the dead.

Gunshot residue patterns on clothing may not be readily visible, particularly if they are on a dark fabric or obscured by blood. In such a case several techniques can be used to visualize the gunshot residue pattern. In the case of patterns obscured by blood, infrared photography can render the pattern visible; blood is virtually transparent to infrared radiation, whereas soot is a strong infrared absorber. Infrared imaging devices have also been successfully used to visualize gunshot residue patterns. Chemical tests may also be used to visualize a gunshot residue pattern on a dark-colored garment or one that is bloodstained. C acid (2-naphthylamine-4,8-disulfonic acid) reacts with traces of nitrates and nitrates in gunshot residue patterns to produce dark red spots. In the Griess test nitrites in the gunshot residue pattern react with the Griess reagent (sulfanilic acid and 2-naphthylamine or sulfanilamide and *N*-(1-naphthyl)ethylenediamine) to produce an azo dye by means of the familiar diazotization reaction. For the C acid and Griess tests the gunshot residue pattern is transferred to a piece of desensitized photographic paper that has been impregnated with the appropriate reagent. The garment is placed on a towel on a laboratory benchtop with the powder pattern uppermost; the impregnated photographic paper is placed with the gelatin-coated side down and covered with a towel. A hot iron is then pressed over the entire pattern to transfer the powder pattern to the photographic paper.

The sodium rhodizonate test is also used in range-

of-fire determinations. Sodium rhodizonate reacts with lead primer residue to produce a blue lead rhodizonate complex. This blue complex can be converted into a scarlet complex by treatment with tartrate buffer (pH 2.8). For the sodium rhodizonate test the pattern is transferred to a sheet of filter paper, which is then sprayed with the sodium rhodizonate reagent, followed by the tartrate buffer solution. Alternatively, the tartrate buffer solution may be saturated with sodium rhodizonate and this saturated solution used to develop the gunshot residue pattern. The sodium rhodizonate test may be made specific for lead by spraying the scarlet pattern produced by the sodium rhodizonate/tartrate buffer combination with dilute hydrochloric acid to produce a blue–violet pattern. Barium ion produces a red complex with sodium rhodizonate; however, the color of this complex is not affected by changing the pH. Mercurous ion, ferrous ion and thiosulfate ion give red or orange colors with sodium rhodizonate solution but these colors fade when treated with dilute hydrochloric acid. Ammunition containing lead-free primers will of course not produce residue containing lead and the sodium rhodizonate test will not produce a color reaction with powder patterns fired with such ammunition. If the ammunition primer contains zinc peroxide zincon can be used to develop a primer residue pattern.

Once the firearms examiner has visualized the powder pattern, he will test-fire the suspect weapon into cloth targets set at different distance from the muzzle of the weapon until a gunshot residue pattern is obtained that closely approximates the questioned pattern in size and density. Care must be taken to use ammunition from the same lot as that believed to have been used to fire the questioned powder pattern. For this reason, police investigators should collect as evidence any unfired ammunition in the possession of the suspect. Failure on the part of defense experts in the John Donald Merrett case to use the same weapon and ammunition from the same lot as that used to produce the questioned powder pattern led to an erroneous opinion as to the range from which the shot that killed Merrett's mother was fired; Merrett was acquitted and went on to commit two other murders. Stippling and soot deposition on the test-fired patterns is readily apparent to the naked eye; however, primer residue patterns on the test-fired targets must be visualized by chemical treatment, using sodium rhodizonate or zincon, as appropriate.

The range from which a shotgun pellet pattern was fired can also be estimated. Although the rule of thumb that a shotgun pellet pattern will spread one inch for every yard the shot mass travels down range may be used for rough estimates, careful range-of-fire determinations require that the firearms examiner have the original pellet pattern, the shotgun believed to have fired it and shotshells from the same lot as that used to fire the questioned pellet pattern. The examiner also needs information about the weather conditions: wind speed and direction obviously effect the size and density of the pellet pattern; ambient temperature also effects the size of the pellet pattern (presumably by changing the rate of combustion of the propellant in the shotshell). The firearms examiner will test fire pellet patterns into paper or cloth targets from a number of different ranges. The examiner may visually compare the test-fired pellet patterns with the questioned pattern and continue his test-firings until he obtains a test-fired pellet pattern that appears to match the questioned pattern in overall size and density. This simple approach can yield surprisingly accurate results. In a blind study ten pellet patterns were fired at randomly selected ranges between 6 feet and 41 feet (2–12.5 m). Visual comparison of these ten patterns with test-fired patterns resulted in range-of-fire estimates with an average absolute error of only 1.5 ft (45 cm) and an average relative error of 6.8%.

Various attempts have been made to introduce statistical methods such as regression analysis into the estimation of the range of fire from a pellet pattern. Statistical methods require some measure of the size of a pellet pattern. If D_{H} is the horizontal spread of the pellet pattern, D_{v} the corresponding vertical spread and N the number of pellets, the following measures of the size of a shotgun pellet pattern can be defined:

$$\langle D\rangle = \frac{(D_{\mathrm{H}} + D_{\mathrm{V}})}{2}$$

$$D = \sqrt{D_{\mathrm{H}}D_{\mathrm{V}}}$$

$$A = D_{\mathrm{H}}D_{\mathrm{V}}$$

$$d = \frac{N}{D_{\mathrm{H}}D_{\mathrm{V}}}$$

If x_i and y_i are the Cartesian coordinates of the ith pellet hole in an arbitrary coordinate system then:

$$S = \sqrt{\sum_{i=1}^{N}[(x_i - \bar{x})^2 + (y_i - \bar{y})^2}$$

where

$$\bar{x} = \sum_{i=1}^{N}\frac{x_i}{N}$$

and

$$\bar{y} = \sum_{i=1}^{N}\frac{y_i}{N}$$

Finally, if R is the radius of the smallest circle that will

just enclose the pellet pattern, R can be measured using a clear plastic overlay marked off with a series of concentric circles. Studies have shown that D_H, D_V, $<D>$, D, S and R are generally linear functions of the range of fire, whereas A and d are not. The calculation of S is laborious and has little to recommend it for practical work.

Regression analysis is familiar to most scientists and a variety of statistical computer software packages are currently available which permit the calculation of regression equations and more importantly the confidence limits for the estimated ranges of fire. The major problem with the application of regression analysis to the problem of range of fire estimation is the large number of test-fired patterns required to determine the regression equation for pellet pattern size versus range of fire. A minimum of 20 test-fired patterns would be required for useful results. A firearms examiner rarely has available this many shotshells of the same lot as that used to fire the questioned pellet pattern. Two alternative approaches have been proposed for the determination of the confidence limits of estimated ranges of fire. In the first, three or more shots are test fired at each of several ranges and the largest and smallest pellet pattern sizes are plotted graphically. The lower confidence limit is the range at which the size of the questioned pellet pattern corresponds to the size of the largest pellet pattern size; the higher confidence limit is the range at which the size of the questioned pellet pattern corresponds to the size of the smallest pellet pattern size. This procedure requires no assumptions about how the shotgun pellets are distributed in the pellet pattern. On the other hand, it is virtually impossible to determine the confidence level of the confidence limits. Confidence limits based on an assumed normal distribution of pellets within the pattern have also been explored. However, experimental evidence indicates that the distribution of pellets within a pattern is unlikely to be Gaussian. For example, 00 buckshot patterns typically consist of a series of superimposed triangles (for a nine-pellet load, two nearly coincident triangles with the third rotated roughly 60° with respect to the first two). No. 2 shot patterns have been shown to have bimodal distributions.

An additional complication in the estimation of the range of fire from a shotgun pellet pattern is the phenomenon called the 'billiard ball effect'. This term refers to the spreading of a shotgun pellet pattern caused by an intermediate target. When shotgun pellets exit a shotgun barrel they initially travel as an elongated mass. If the pellet mass encounters an intermediate target the leading pellets will be slowed down; the trailing pellets will overtake the leading pellets, colliding with them and thus causing them to fly off at eccentric angles. If the pellets go through the intermediate target to strike the final target, the resulting pellet pattern may be larger than would be produced at the same muzzle-to-final-target distance in the absence of the intermediate target. In other words, the pellet pattern in the final target may appear to have been fired from a greater distance than was actually the case. The larger (and hence more massive) the pellets the denser the intermediate target must be in order for it to produce a billiard ball effect. Metal or plastic window screen will produce a significant billiard ball effect with no. 2 birdshot, but the effect of these materials on 00 buckshot is trivial. Human skin can act as an intermediate target; consequently, care should be taken in making estimates of ranges of fire from radiographic images. As the muzzle-to-intermediate-target distance increases the billiard ball effect diminishes and eventually disappears altogether. As the pellets travel away from the muzzle the pellets in the shot mass separate from one another; eventually each pellet is traveling on its own trajectory and the slowing of the leading pellets by the intermediate target will not result in collisions between pellets. The firearms examiner may be alerted to the existence of an intermediate target by the presence of markings (e.g. from window screen) or trace evidence (e.g. paint or wood splinters) on the shotgun pellets.

See also: **Firearms**: Weapons, Ammunitions and Penetration; Residues; Laboratory Analysis. **Pattern Evidence**: Shotgun Ammunition on a Target.

Further Reading

Abrink A, Andersson C and Maehly AC (1984) A video system for the visualization of gunpowder patterns. *Journal of Forensic Sciences* **29**:1223–1224.

Barnes FC and Helson RA (1974) An empirical study of gunpowder residue patterns. *Journal of Forensic Sciences* **19**:448–462.

Bartsch MR, Kobus HJ and Wainwright KP (1997) Update on the use of the sodium rhodizonate test for the detection of lead originating from firearm discharges. *Journal of Forensic Sciences* **41**:1046–1051.

Beijer R (1994) Experience with zincon, a useful reagent for the determination of firing range with respect to leadfree ammunition. *Journal of Forensic Sciences* **39**:981–987.

Bhattacharyya C and Sengupta PK (1989) Shotgun pellet dispersion in a Maxwellian model. *Forensic Science International* **41**:205–217.

Boyer DA, Marshall LD, Trzicak LM and Rowe WF (1989) Experimental evaluation of the distribution of the pellets in shotgun pellet patterns. *Forensic Science International* **42**:51–59.

Breitenecker R and Senior W (1967) Shotgun patterns I. An experimental study on the influence of intermediate targets. *Journal of Forensic Sciences* **12**:193–204.

Fann CH, Ritter WA, Watts RH and Rowe WF (1986) Regression analysis applied to shotgun range-of-fire estimations: results of a blind study. *Journal of Forensic Sciences* **31**:840–854.

Feigl F and Anger V (1972) *Spot Tests in Inorganic Analysis*. New York: Elsevier.

Fisher RS (1993) Shotgun wounds. In: Spitz WU (ed.) *Spitz and Fisher's Medicolegal Investigation of Death: Guidelines for the Application of Pathology to Crime Investigation*, 3rd edn. Springfield: Charles C Thomas.

Jauhari M (1973) Functional relationships between the range of firing and the various characteristics of a pellet pattern. *Journal of the Indian Academy of Forensic Sciences* **12**:5–9.

Jauhari M, Chatterjee SM and Ghosh PK (1972) Statistical treatment of pellet dispersion data for estimating range of firing. *Journal of Forensic Sciences* **17**:141–149.

Jauhari M, Chatterjee SM and Ghosh PK (1974) A comparative study of six parameters for estimating range of firing from pellet dispersion. *Journal of the Indian Academy of Forensic Sciences* **13**:17–24.

Mattoo BN and Nabar BS (1969) Evaluation of effective shot dispersion in buckshot patterns. *Journal of Forensic Sciences* **14**:263–269.

Menzies RC, Scroggie RJ and Labowitz DI (1981) Characteristics of silenced firearms and their wounding effects. *Journal of Forensic Sciences* **26**:239–262.

Moreau TS, Nickels ML, Wray JL, Bottemiller KW and Rowe WF (1985) Pellet patterns fired by sawed-off shotguns. *Journal of Forensic Sciences* **30**:137–149.

Nag NK and Lahiri A (1986) An evaluation of distribution of pellets due to firearm discharge. *Forensic Science International* **32**:151–159.

Rowe WF (1988) Firearms identification. In: Saferstein R (ed.) *Forensic Science Handbook*, vol. II. Englewood Cliffs: Prentice Hall.

Rowe WF (1991) Statistics in forensic ballistics. In: Aitken CGG and Stoney DA (eds) *The Use of Statistics in Forensic Science*. Chichester: Ellis Horwood.

Rowe WF and Hanson SR (1985) Range-of-fire estimates from regression analysis applied to the spreads of shotgun pellet patterns: results of a blind study. *Forensic Science International* **28**:239–250.

Spitz WU (1993) Gunshot wounds. In: Spitz WU (ed.) *Spitz and Fisher's Medicolegal Investigation of Death: Guidelines for the Application of Pathology to Crime Investigation*, 3rd edn. Springfield: Charles C Thomas.

Steinberg M, Leist Y, Goldschmidt P and Tassa M (1984) Spectrophotometric determination of nitrites in gunpowder residue on shooter's hands. *Journal of Forensic Sciences* **29**:464–470.

Stone IC and Petty CS (1974) Examination of gunshot residues. *Journal of Forensic Sciences* **19**:784–788.

Walker JT (1940) Bullet holes and chemical residues in shooting cases. *Journal of Criminal Law, Criminology and Police Science* **31**:497–521.

Wray JL, McNeil JE and Rowe WF (1983) Comparison of methods for estimating range of fire based on the spread of buckshot patterns. *Journal of Forensic Sciences* **28**:846–857.

Residues

W F Rowe, The George Washington University, Washington DC, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0535

Gunshot residue consists of a variety of materials: particles of the projectile or the projectile's jacket, unburnt particles of smokeless powder, partially burnt particles of powder, combustion products and particles of primer residue. These materials are projected from the muzzle of the weapon in a conical cloud. The particles are slowed down by air resistance, with the larger particles traveling the greater distances. Gunshot residue may also escape from various openings in the firearm: from the space between the chamber and the barrel in revolvers, from the ejection ports of self-loading or automatic firearms and even from the trigger hole. Gunshot residue escaping in this way may be deposited on the hands of the shooter and on the hands of someone who is grappling for possession of the firearm at the instant of its discharge. In the field of forensic science, the detection of gunshot residue becomes important in two aspects of the investigation of a shooting incident: the determination of the range from which a shot was fired and as an aid in the identification of the shooter. The first involves the examination of powder patterns on the surfaces of targets, and the second involves the analysis of residues removed from the hands of a suspected shooter.

The gross appearance of a gunshot wound can provide some insight as to the range from which it was fired. Distant gunshot wounds are fired from a sufficient distance that no powder residue reaches the body surface; only the bullet reaches the target. The injury in this case typically consists of a circular or elliptical defect surrounded by a contusion ring (a ring of bruising caused by the bullet's stretching and tearing of the skin); the contusion ring may be overlain by a gray ring, a ring of material wiped from the surface of the bullet as it passed through the skin. Bullet wipe consists of powder residue, bullet lubricant and traces of metal and metal oxides from the surface of the bullet. A close range gunshot wound is one that is inflicted from a sufficiently short range that powder residue reaches the body surface. A close range gunshot wound will have the same features as a distant shot plus deposition of tattooing (also called stippling) and soot (finely divided combustion products from the burning of propellant). Tattooing or stippling consists of particles of unburned and partially burned propellant that are driven into the skin surface or the surface of the shooting victim's clothing. As the

range of fire shortens the pattern of tattooing and soot deposition becomes both smaller and denser. At near-contact range (less than about 1 in (2.5 cm) the hot gases comprising the muzzle flash will singe the target surface, burning hair and natural textile fibers and melting synthetic fibers. The muzzle flash can split woven fabrics along the warp and weft directions. Forensic pathologists distinguish between loose contact and tight contact gunshot wounds. Loose contact gunshot wounds may show gunshot residue blown along the body surface or between layers of clothing. In addition to singeing from the muzzle flash, loose contact gunshot wounds may show ironing, i.e. a stiffening of fabric due to contact with a hot surface, caused by the weapon's muzzle. Tight contact wounds usually show no gunshot residue on the body surface. If a tight contact gunshot wound is inflicted over a boney plate such as the vault of the skull or the sternum, it may appear as a stellate (starlike) defect with a patterned contusion from the weapon's muzzle. In this case, the hot gases issuing from the muzzle of the weapon separate the soft tissue from the underlying bone. The resulting pocket of hot gases pushes the soft tissue against the weapon's muzzle; if gas pressure is high enough the soft tissue may tear, allowing the gas to escape. Blood may be blown back on to the hand or arm of the shooter; blood and tissue may also be blown into the muzzle of the weapon.

The determination of the range of fire from a powder pattern requires four things: the original powder pattern; the weapon that produced the pattern; ammunition from the same lot as that used to produce the original pattern; and knowledge of the weather conditions. If the original powder pattern is on clothing its preservation is straightforward. On the other hand, if the powder pattern is on the skin of a shooting victim, a scaled photograph should be taken using color film. Because skin blemishes (small moles, freckles and the like) may be mistaken for tattooing or stippling – even in a color photograph – the examiner who will make the range of fire determination should be able to examine the powder pattern himself. From time to time recommendations for the excision of powder patterns from the bodies of deceased have appeared in the forensic science literature. The skin is sutured to a metal hoop and then carefully cut away from the body. The excised skin is then treated with a preservative such as formaldehyde. As an alternative, the excised tissue may be preserved by freezing. Although excision of the powder pattern may be necessary for laboratory analysis, the firearms examiner needs to be aware that the actual removal of the powder pattern is fraught with problems. To begin with, the skin may be stretched or it may shrink, altering the size of the powder pattern. At trial, the judge may exclude the admission of the powder pattern as evidence on the ground that the prejudicial effect of the admission of such evidence outweighs its probative value. Finally, relatives of the deceased may regard the excision of the powder pattern as a desecration of the victim's body.

A questioned powder pattern on clothing or other textile may require special treatment to render it visible, particularly if the fabric is dark colored or soaked with blood. The soot in a powder pattern strongly absorbs infrared radiation. Therefore, it may be possible to photograph a powder pattern on the surface of a dark garment using infrared film and special camera filters that exclude visible light. The hemoglobin in blood does not strongly absorb infrared radiation; therefore, infrared photography can be used to visualize powder patterns on bloodstained garments. Forensic light sources (e.g. Polilight® or Omnichrome® 9000) have also proved useful for the visualization of powder patterns: partially burnt nitrocellulose particles fluoresce when illuminated with 415 nm light. Many forensic science laboratories use chemical methods to visualize powder patterns. First, the powder pattern is sprayed with a dilute solution of sodium rhodizonate, which reacts with the residues of lead primer compounds to form a blue lead rhodizonate complex; the color of the lead complex is changed from blue to red by spraying with an aqueous tartrate buffer solution. The primer residue pattern is photographed before additional tests are carried out.

Users of the sodium rhodizonate test have noted several problems with it. An aqueous solution of sodium rhodizonate decomposes very rapidly; and the final blue color may be subject to unpredictable rapid fading. A study of the chemistry of the reaction of sodium rhodizonate with lead has also revealed that if the complexation reaction takes place at neutral pH rather than at pH 2.8 (the pH of the tartrate buffer) the formation of tetrahyroquinone will be favored over the formation of the scarlet lead–rhodizonate complex. Tartrate ion seems to participate directly in some way in the complexation reaction. To deal with these problems a modification of the sodium rhodizonate test has been proposed. In this new procedure the powder residue is transferred to a sheet of filter paper. The filter paper is then sprayed first with tartrate buffer then with a saturated aqueous sodium rhodizonate solution or with a tartrate-buffered saturated sodium rhodizonate solution (which has a half-life at room temperature of about ten hours). If lead is present the scarlet lead–rhodizonate complex is formed. The scarlet complex is decomposed by spraying the pattern with 5% aqueous hydrochloric acid until the blue color reaches maximum intensity. The hydrochloric acid is removed

by drying the filter paper with a hair drier. The blue pattern will be indefinitely stable.

After the sodium rhodizonate test the combustion products of smokeless powder are visualized with the Griess test. A sheet of unexposed photographic paper that has been treated with developer and fixer and carefully washed is impregnated with Griess reagent. The questioned powder pattern is placed over the impregnated photographic paper in contact with the gelatin-coated surface and covered with a dampened towel. An electric clothes iron is used to iron the powder pattern on to the photographic paper where nitrites in the pattern react with the Griess reagent to produce a rose-colored azo dye through a diazo coupling reaction. The reaction of the powder pattern with Griess reagent can be enhanced by spraying the pattern with aqueous sodium hydroxide solution and heating it in a laboratory oven. Under alkaline conditions nitrocellulose disproportionates to yield nitrite ions. The interpretation of the powder patterns visualized by the Griess reaction can be complicated by fabric dyes with similar colors to the Griess diazo reaction product: these dyes may leach out of fibers and bleed into the gelatin layer of the photographic paper.

Bloodstains can interfere with the Griess test. Traces of blood can be transferred to the test paper, either partially or completely obscuring the color of the reaction product. The Maiti test was developed to avoid this masking problem. In the Maiti test sheets of photographic paper are prepared by fixation in sodium thiosulfate solution, washed and then dried. Then the sheets are soaked in a solution of *p*-nitroaniline, α-naphthol and magnesium sulfate (0.25% of each in 1:1 aqueous alcohol). The impregnated paper is now dried and preserved for later use. When a powder pattern is to be visualized, a sheet of impregnated photographic paper is placed on the laboratory bench emulsion side up. The item of clothing bearing the powder pattern is then placed on top of the photographic paper with the surface bearing the powder pattern placed in contact with the emulsion layer of the photographic paper. A cloth moistened with 10% acetic acid is placed on top of the item of clothing and the whole stack is pressed with a hot iron to transfer powder particles to the photographic paper. Finally, the surface of the emulsion is swabbed with a 10% sodium hydroxide solution. The resulting powder pattern consists of blue flecks on a pale yellow background.

The weapon that produced the original powder pattern is test fired into pieces of white cotton sheeting at various ranges until a test-fired pattern with the same size and density as the original is produced. The original weapon must be used because its condition (e.g. degree of erosion of the bore) can affect the powder pattern. Likewise ammunition from the same lot as that used to produce the original powder pattern must also be used to obtain the test-fired patterns. Different lots of the same brand of ammunition may contain smokeless powder from different lots of propellant. Ammunition manufacturers often change lot numbers when a sublot of one component (projectile, casing, propellant or primer) has been exhausted and the use of a new sublot of that component is begun. Propellants from different sublots may differ in the quantity of gunshot residue that they produce. To insure that the firearms examiners have ammunition from the same lot as that used to fire the original powder pattern, investigators should seize as evidence any remaining boxes of ammunition in the suspect's possession.

Knowledge of the weather conditions at the time of the shooting is essential to the estimation of the range from which a powder pattern was fired. Wind and rain will disperse the gunshot residue. The ambient temperature may also affect the burning rate of the propellant and thus influence the appearance of the powder pattern.

The range from which a shotgun pellet pattern was fired can be estimated from the pattern of deposition of gunshot residue as discussed above or from the size and density of the pellet pattern. The making of such range estimates from pellet patterns is complicated by the existence of different shotgun chokes. Many shotgun barrels are produced with constrictions at the muzzle whose purpose it is to concentrate the shot pattern. Common designations of the degree of choke of a shotgun barrel are the following: cylinder-bore (no choke), improved cylinder (slight choke), modified choke and full choke. The choke of a shotgun barrel can be determined by firing pellet patterns at a range of 40 yards (37 m): a full-choke barrel will place 65–75% of the pellets within a 30-inch (75 cm) circle; a modified choke barrel will place 45–65% within the circle; an improved cylinder barrel will place 35–45% within the circle; and the cylinder bore barrel will place 25–35% within the circle. The two barrels of a double-barreled shotgun frequently have different chokes. The choke of a shotgun can also be changed with barrel inserts or adjustable compensators. Obviously the choke of a shotgun affects the size and density of the pellet pattern it fires at any given range. Several approaches have been used to estimate the range of fire from the size and density of a pellet pattern. Some firearms examiners have used the rule of thumb that a shotgun pellet pattern spreads about one inch (2.5 cm) for each yard (meter) that the shot string travels down range. Another approach is to test fire the weapon that fired the questioned pellet pattern at

various ranges into paper or cardboard targets until a pellet pattern of the same size and shot density as that of the questioned pellet pattern is obtained. The test-fired patterns must be obtained using shotshells from the same lot as that used to fire the questioned pattern. Studies have shown that the sizes of shotgun pellet patterns produced by different lots of the same brand of ammunition are statistically different. Knowledge of the weather conditions at the time the questioned pattern was fired is also important. The ambient temperature has been shown to affect the spread of shotgun pellets. Attempts have also been made to apply regression analysis to the estimation of the range of fire from the size of a pellet pattern. However, the application of regression analysis to the estimation of the range of fire of a shotgun pellet pattern requires that a large number of test-fired pellet patterns (more than twenty if the confidence limits of the estimate are to be forensically useful). Rarely will sufficient shotshells from the same batch as that used to fire the questioned pattern be available to the firearms examiner for use to be made of regression analysis.

Although many of the same procedures may be used both to visualize powder patterns and to detect gunshot residue on the hands of a suspected shooter, it must be strongly emphasized that the purposes of these two types of analysis are quite different. In the first case, the test results are used to estimate the range from which a gunshot was fired, and in the second case the test results are used to link a suspect with the discharge of a firearm. The first test for the presence of gunshot residue on the hands of a suspect was the dermal nitrate test (also called the paraffin test, the Gonzalez test or the diphenylamine test) which was introduced in the 1930s. The test was developed by Tomas Gonzalez, chief of police of Mexico City and later Chief Medical Examiner of New York City. Gunshot residue was removed from the suspect's hand using a paraffin glove. A layer of melted paraffin was carefully 'painted' on to the hand and then while it was still soft the paraffin was reinforced with a layer of surgical gauze. Further layers of melted paraffin and gauze would be applied to produce a thick glove that could be handled without danger of disintegration. The glove would be allowed to cool and then it would be cut from the hand. Drops of diphenylamine reagent (typical formula: 0.25 g diphenylamine and 0.25 g *N,N'*-diphenylbenzidine in 70 ml concentrated sulfuric acid) would be applied to the interior surface of the glove. Gunshot residue would be revealed by the presence of blue flecks whose blue color streamed off into the reagent solution. The diphenylamine reagent produces a blue color when it reacts with partially burnt particles of smokeless powder. The diphenylamine reagent also reacts with unburned and partially burned particles of smokeless powder. Because the dermal nitrate test is actually a test for oxidizing agents a variety of materials may interfere with it: household bleaches (sodium and calcium hypochlorites); water treatment chemicals (calcium hypochlorite); fertilizers (ammonium nitrate); and explosives. Even the nitrates in urine could produce a false positive test. Cosmetics and tobacco have also been found to interfere with the test. Because of these problems with the dermal nitrate test an Interpol seminar unanimously recommended in 1963 that the dermal nitrate test no longer be used for either investigative or evidentiary purposes.

In 1959 Harrison and Gilroy published a chromogenic analysis scheme for the detection of primer residues. The suspect's hand were first swabbed with a small square of clean cotton cloth which had been dampened with 0.1 N hydrochloric acid. A drop of a 10% alcoholic solution of methyltriphenylarsonium iodide was placed on the cloth. An orange ring indicated the presence of antimony. After drying, the cloth was next tested with two drops of 5% aqueous sodium rhodizonate solution. A red color developing within the orange ring indicated the presence of barium or lead or both. The cloth swab was again dried and then one or two drops of 1:20 hydrochloric acid were added to the red colored area. A blue color indicated the presence of lead; if the red color remained inside the blue-colored ring barium was also present. The addition of the 1:20 hydrochloric acid was important because mercurous ion, ferrous ion and thiosulfate ion give a red or orange color with sodium rhodizonate, but hydrochloric acid destroys the colored complexes with these ions. The detection limits for the three metal ions were determined by Harrison and Gilroy to be 4 μg of antimony in the presence of 1.5 mg of lead and 10 mg of barium, 10 μg of barium in the presence of 1.5 mg of lead (higher quantities of lead masking the color of the barium rhodizonate complex) and 2.5 μg of lead.

Instrumental methods of analysis have replaced the chromogenic methods discussed above in modern crime laboratories. Atomic absorption (AA) spectrophotometry has become the method of choice for many forensic science laboratories for the analysis of suspected gunshot residue. AA is much less expensive than neutron activation analysis (NAA), (see below) and it can detect lead, as well as antimony and barium. The detection limits of AA for these elements are higher than those of NAA but are adequate for most gunshot residue samples. A typical AA spectrophotometer consists of a light source, a monochromator, a sample atomizer, a photomultiplier detector and some type of read-out device. The light sources used in AA are hollow cathode lamps, which

consist of glass envelopes with windows through which the radiation generated at the hollow cathode passes. The hollow cathode is a cylinder made of or coated with the element for whose analysis the lamp is to be used. The lamp is filled with a low pressure of argon gas. When a high voltage (5000 V) is applied between the hollow cathode lamp's anode and cathode the argon fill gas is ionized. The positively charged argon ions are accelerated toward the hollow cathode whose surface they bombard. Atoms are knocked from the surface of the hollow cathode in excited electronic states. The excited atoms emit the wavelengths of light characteristic of the element comprising or coating the hollow cathode. The light emitted by the hollow cathode lamp passes through a monochromator which isolates the wavelength of light which has been selected as the analytical wavelength. The emission profile of the analytical wavelength is much narrower than the absorption profile of gas phase atoms at the same wavelength. This is a necessary requirement for quantitative elemental analysis by AA because Beer's Law is valid only when the absorptivity (or extinction coefficient) of the absorbing species is constant over the range of wavelengths passing through the sample.

The analytical wavelength next passes through the sample atomization region of the spectrophotometer where the analytical sample is vaporized and atomized. The analytical wavelength passes through the resulting cloud of gas phase atoms. Flame atomization has been a mainstay of the AA technique. The analytical sample (in the form of a liquid) is aspirated into a gas burner (burning acetylene in oxygen or nitrous oxide) where the heat of the flame first vaporizes it and then dissociates it into its component atoms. Flame AA is not appropriate for the analysis of gunshot residue. Lead, antimony and barium all form refractory oxides in the flame and so are unavailable to absorb the analytical wavelength. The use of carbon rod atomizers has also been explored; however, it was found that barium could not be determined by this method because it forms refractory barium carbide. Tantalum strip atomizers and graphite furnaces have been successfully used as atomizers for the analysis of gunshot residue.

AA is compatible with a wide variety of sampling techniques. Gunshot residue can be removed from the hands by swabbing or by dipping them into plastic bags containing a dilute acid solution. Tape lifts and film lifts can also be used to collect gunshot residue samples for AA analysis. The lifts must be ashed in an oxygen plasma; the primer residues are dissolved in an acid solution for AA analysis. Regardless of the method used to collect the gunshot residue, control samples of the materials used in the collection (e.g. acid solution, swab, lifting tape or lifting film) must also be submitted for analysis to verify that these materials do not contain the elements of interest.

Scanning electron microscopy (SEM) has been used by some forensic science laboratories for the detection of gunshot residue. In the SEM, an electron beam is swept raster-fashion over the surface of the specimen. A detector collects electrons being produced by the surface of the specimens; an image is formed by converting the detector current into the intensity of a spot on a cathode ray tube whose scan is synchronized with the scan of the electron beam over the specimen surface. Two types of electrons are commonly detected: secondary electrons (electrons 'knocked' off the surface of the specimen by the electron beam) and backscattered electrons (electrons from the electron beam that are deflected backwards by collisions with the electron clouds surrounding atoms in the specimen). Backscattered electron detectors have proven to be particularly useful in the analysis of primer residues because heavy atoms such as lead, antimony and barium are very efficient backscatterers. Particles containing these elements appear brighter on the SEM image display. This allows the SEM operator to pick out the primer particles more quickly. Primer particles take a number of forms; they may be single spheres or clusters of globules. The majority of primer residue particles are spheres ranging from 0.5 μm to 5.0 μm in diameter.

The morphology of primer particles is highly characteristic; however, the elemental makeup of the particles must also be examined. If the primer residue particles are bombarded with high-energy electrons they can be made to produce X-rays. The high-energy electrons penetrate the electron shells of atoms at the surface of the primer residue particle. Collisions with inner shell electrons can lead to removal of an inner shell electron from the atom, producing a hole in the K shell. This hole can be filled by an outer electron falling into it; when it falls, the outer electron loses energy in the form of an X-ray photon. The emitted X-ray photon is detected by an energy-dispersive X-ray analyzer (hence the name SEM/EDX). The X-ray emission lines of antimony, barium and lead are given in **Table 1**.

In 1994 the ASTM (American Society for Testing and Materials) Committee E-30 on Forensic Science adopted guidelines (Standard E 1588) for SEM/EDX analysis of gunshot residue, including guidelines for the interpretation of the results of the elemental analysis. The recommended operating parameters for the SEM/EDX system were: (1) that the SEM when operating in the backscattered mode be capable of detecting potential gunshot residue particles down to a diameter of 0.5 μm; (2) that the instrument be

Table 1 X-ray emission lines (KeV)

No.	*Element*	$K\alpha_1$	$K\alpha_2$	$K\beta_1$	$L\alpha_1$	$L\alpha_2$	$L\beta_1$	$L\beta_2$	$L\gamma_1$
51	Sb	26.3591	26.1108	29.7256	3.60472	3.59532	3.84357	4.10078	4.34779
56	Ba	32.1936	31.8171	36.3782	4.46626	4.45090	4.82753	5.1565	5.5311
82	Pb	74.9694	72.8042	84.936	10.5515	10.4495	12.6137	12.6226	14.7644

capable of producing a 3:1 signal-to-noise ratio for the lead $L\alpha$ emission line from a lead particle no larger than 1 μm in diameter; and (3) that the instrument be capable of resolving the $L\alpha_1$, $L\beta_1$ and $L\beta_2$ emission lines. To fulfill these requirements the SEM/EDX must be capable of operating at a 20 KeV accelerating potential or higher.

The following combinations of elements have been observed only in primer residue particles: lead, antimony and barium; and antimony and barium. Other combinations are consistent with primer residue but might derive from other sources: barium, calcium, silicon with a trace of sulfur; lead and antimony; lead and barium; lead; barium.

Samples for SEM analysis may be collected in a variety of ways. Both tape lifts and film lifts have been used. The lifts are sputter-coated with carbon to insure electrical conductivity and prevent charging of the tape or film surfaces. Metal SEM sample stubs coated with adhesive can also be used. Carbon is used to sputter-coat the samples because its X-ray emissions are blocked by the beryllium window of the SEM's X-ray analyzer. Because observation of the morphology of the particles is essential to the identification of primer residue, sample collection methods (e.g. swabbing with dilute acid or dipping in dilute acid) in which the primer residue particles are dissolved cannot be used.

Some forensic science laboratories have begun to use X-ray microfluorescence for the analysis of gunshot residue. In this technique a narrow beam of X-rays is focused on the sample. The X-rays ionize some of the atoms in the sample by removing inner shell electrons; outer shell electrons fall into the resulting vacancies, emitting fluorescent X-rays as they do so. The emitted X-rays are then analyzed by an energy-dispersive X-ray spectrometer. Instruments such as the Kevex Omnicron energy dispersive X-ray microfluorescence spectrometer are capable of scanning a large area and generating maps of the intensities of the fluorescent X-rays. Consequently, X-ray microfluorescence has the capability not only of determining the presence of lead, antimony and barium in primer residue but also of showing their distribution pattern on the target surface.

The use of neutron activation analysis (NAA) for the detection of primer residues was developed during the heyday of the exploitation of NAA in forensic science. In NAA the sample is exposed to bombardment by thermal neutrons in the core of a nuclear reactor. Stable isotopes of lead, antimony and barium (**Table 2**) capture neutrons and are converted into radioactive isotopes. After the sample is removed from the reactor the induced radioactivity is measured with a γ-ray spectrometer. The γ-ray energies measured identify the radioactive isotopes present in the sample (**Table 3**) and the number of γ-rays indicates the number of radioactive nuclei present. Although NAA has a number of attributes that commend it for the detection of primer residues, such as very low detection limits and compatibility with a number of

Table 2 Stable isotopes of barium and antimony

Isotope	*% Natural abundance*
^{134}Ba	2.42
^{135}Ba	6.59
^{136}Ba	7.81
^{137}Ba	11.32
^{138}Ba	71.66
^{121}Sb	57.25
^{123}Sb	42.75

Table 3 Radioactive isotopes of barium, sodium and antimony

Radioactive isotope	*Half-life*	*γ-Ray emissions (MeV)*
^{139}Ba	82.9 minutes	0.166
		1.270
		1.430
^{24}Na	14.96 hours	1.369 (100%)
		2.754 (100%)
^{122}Sb	2.80 days	0.564 (66%)
		1.14 (1%)
		1.26 (1%)
^{124}Sb	60.4 days	0.603 (97%)
		0.644 (7%)
		0.72 (14%)
		0.967 (2%)
		1.048 (2%)
		1.31 (3%)
		1.37 (5%)
		1.45 (2%)
		1.692 (50%)
		2.088 (7%)

sampling methods, it also suffers from a number of flaws. The most obvious is the cost of this method of analysis. Another is the procedure's inability to detect lead. Although the common stable isotopes of lead may be converted into radioactive isotopes by neutron bombardment, none of these isotopes emits γ-rays. Low-resolution γ-ray detectors have also posed problems. When such detectors are used the most intense γ-ray emission of ^{139}Ba can be obscured by a γ-ray emission of ^{24}Na. Perspiration will produce very high levels of sodium in samples taken from the hands. Initially, the interference from sodium was removed by radiochemical separation. However, Krishnan has developed a NAA procedure using a series of irradiations followed by cool down periods to alleviate the interference from ^{24}Na. For example, the sample is irradiated at a neutron flux of 5×10^{12} neutrons cm^{-2} sec^{-1} for 40 mins. The induced barium radioactivity is counted immediately. The sample is returned to the nuclear reactor for irradiation at the same flux for five hours. After a two-day cool down period the induced antimony radioactivity is counted. The long cool-down period permits most of the ^{24}Na radioactivity to disappear.

Samples for NAA analysis can be obtained in a variety of ways. The paraffin glove method used in the dermal nitrate test was the first to be used. Later it was deemed easier to remove the primer residues by dipping the hands into a plastic bag containing a dilute nitric acid solution or by swabbing the hands with cotton-tipped swabs dampened with a dilute acid solution. The swabbing method of sample collection is used in commercially packaged NAA gunshot residue collection kits. Swabbing is now the collection method of choice. Swabs can be collected from several areas on the firing and non-firing hands (e.g. web between thumb and forefinger, back of the hand and palm of the hand). High levels of antimony and barium in the web area of the right hand of a suspect with low levels elsewhere would be consistent with the suspect having discharged a firearm. On the other hand, elevated levels of antimony and barium on the palm of one or both hands would be consistent with the suspect merely handling a firearm. Regardless of the method of sample collection chosen, the low detection limits of NAA require that negative control samples (reagent solutions, swabs and the like) also be submitted for analysis so that the absence of antimony and barium from the sampling media can be demonstrated. Early in the development of the swabbing technique some samples of cotton swabs were found to have high levels of barium in their wooden shafts.

A 1992 survey of forensic science laboratories in the United States and in two Canadian provinces found that 44% of the respondent laboratories used AA by itself. About 26% of the laboratories used SEM/EDX by itself and 29% combined AA and SEM/EDX. Only 2% of the respondent laboratories were still using NAA for gunshot residue detection. The laboratories reported a wide range of threshold values for reporting a positive result: for Pb 0.1–2.0 μg ml^{-1}; for Ba 0.1–1.0 μg ml^{-1}; and for Sb 0.02–0.2 μg ml^{-1}.

Ammunition manufacturers have begun to introduce lead-free primers. Not only do these primers lack the lead azide or lead styphnate primary high explosives they also do not contain antimony sulfide or barium nitrate. The primary high explosives used in lead-free primers include diazodinitrophenol and tetracene. Zinc peroxide or strontium nitrate may be added as oxidizers and fine titanium particles may also be found in the primer mixture. Elemental analysis of primer residues produced by lead-free primers reveals particles containing zinc and titanium or particles containing strontium. Lead and antimony from the surfaces of lead-alloy bullets may also be incorporated in the primer residue particles. Firearms which have been previously used to fire ammunition having lead-based primers have been found to produce residue particles containing up to four elements. Primer residues produced by lead-free primers have similar morphologies to those produced by conventional primers: spheres and clusters of spheres. The mechanism by which primer residue particles are formed is basically the same: the metals condense from a hot vapor. The SEM/EDX therefore remains a viable method for identifying primer residues. However, residues from fireworks and flares have been found to contain particles with a similar morphology to those produced by primers containing strontium.

A growing body of research has been focused on the analysis of the organic constituents of smokeless powder, which may also be detectable in gunshot residue. Smokeless powder consists of nitrocellulose to which a variety of compounds may have been added: the energetic plasticizers nitroglycerin and dinitrotoluene, stabilizers such as diphenylamine and ethyl centralite, and nonenergetic plasticizers such as dibutyl phthalate and triacetin. Gas chromatography–mass spectrometry (GC-MS) and high-performance liquid chromatography (HPLC) have both been applied to the detection of the organic additives in smokeless powders; however, micellar electrokinetic capillary electrophoresis (MECE) has shown the greatest promise as a method for the detection of organic compounds in gunshot residue. In conventional capillary electrophoresis ionic species are separated on the basis of their electrophoretic mobilities (which are dependent on such factors as the effective sizes of the ions, their electrical charges and the viscosity of the medium in which the electrophoresis is

taking place). If the wall of the capillary bears ionized groups (e.g. Si-OH in the case of fused silica capillaries) ions of opposite charge will be attracted to the wall of the capillary, leaving the solution in the capillary with a net electrical charge. Application of a voltage through the capillary results in a net flow of the solution through the capillary. This phenomenon is called electroendosmosis (EEO). In MECE an anionic detergent such as sodium docecyl sulfate (SDS) is added to a buffer solution; the detergent forms micelles, spherical aggregates in which the nonpolar tails of the detergent molecules project into the center of the aggregates and their electrically charged heads project out into the buffer solution. The anionic detergent micelles have a large net negative charge, which gives them a large anodic electrophoretic mobility. If the electrophoresis buffer has a large EEO flow toward the cathode (as most do) the capillary will contain a fast moving aqueous phase and a slow-moving micellar phase. Mixtures of organic compounds can be separated based on the compound's differential solubility in the micelles: compounds that are very soluble in the micelles migrate with the slow-moving micelles, whereas compounds that are more soluble in the buffer will migrate more rapidly. The eluted compounds are detected by their absorption of ultraviolet light. The advantage of MECE over GCMS is that MECE separations take place at room temperature so that temperature labile compounds (e.g. nitroglycerin) are less likely to undergo decomposition. Because of its flow profile MECE is capable of higher resolution than HPLC.

In some cases criminal defendants claim that a negative test for the presence of gunshot residue on their hands shows that they did not discharge a firearm. The interpretation of the results of tests for gunshot residues is not so simple. First of all, many firearms deposit little or no detectable residue on the shooter's hands. Surveys of firearm suicides in which handswabs were collected and submitted for gunshot residue have found gunshot residue detection rates which varied from 62% to 38%. In the case of a living shooter the gunshot residue may be removed by washing the hands; it may also be rubbed off the hands on to clothing. Because of the possibility that gunshot residue may be deliberately removed or inadvertently lost from a shooter's hands other sources of gunshot residue should be considered. Gunshot residue may be deposited on the face and hair of the shooter or on his clothing. Gunshot residue deposited in these areas will generally be retained longer than gunshot residue on the hands. Nasal mucus has also been explored as a source of gunshot residue samples. The problem with these alternate sources of gunshot residue is that finding traces of gunshot residue on the face, in the hair, on clothing or in nasal mucus of a criminal suspect only establishes that the suspect was near a firearm when it was discharged.

See also: **Analytical Techniques**: Microscopy. **Firearms**: Weapons, Ammunitions and Penetration; Range; Residues; Laboratory Analysis. **Pattern Evidence**: Shotgun Ammunition on a Target.

Further Reading

American Society for Testing and Materials (1999) Standard guide for gunshot residue analysis by scanning electron microscopy/energy-dispersive spectroscopy. *Annual Book of ASTM Standards*, vol. 14.02. West Conshohocken, Pennsylvania: American Society for Testing and Materials.

Andrasko J and Maehly AC (1977) Detection of gunshot residues on hands by scanning electron microscopy. *Journal of Forensic Sciences* **22**:279–287.

Barnes FC and Helson RA (1974) An empirical study of gunpowder residue patterns. *Journal of Forensic Sciences* **19**:448–462.

Bartsch MR, Kobus HJ and Wainwright KP (1997) Update on the use of the sodium rhozidonate test for the detection of lead originating from firearm discharges. *Journal of Forensic Sciences* **41**:1046–1051.

Beijer R (1994) Experience with zincon, a useful reagent for the determination of firing range with respect to leadfree ammunition. *Journal of Forensic Sciences* **39**:981–987.

Brazeau J and Wong RK (1997) Analysis of gunshot residue on human tissues and clothing using x-ray microfluorescence. *Journal of Forensic Sciences* **42**:424–428.

Breitenecker R and Senior W (1967) Shotgun Patterns I. An experimental study on the influence of intermediate targets. *Journal of Forensic Sciences* **12**:193–204.

DeGaetano D and Siegel JA (1990) Survey of gunshot residue analyses in forensic science laboratories. *Journal of Forensic Sciences* **35**:1087–1095.

DeGaetano D, Siegel JA and Klomparens KL (1992) Comparison of three techniques developed for sampling and analysis of gunshot residue by scanning electron microscopy/energy-dispersive x-ray analysis (SEM/EDX). *Journal of Forensic Sciences* **37**:281–300.

Fann CH, Ritter WA, Watts RH and Rowe WF (1986) Regression analysis applied to shotgun range-of-fire estimations: results of a blind study. *Journal of Forensic Sciences* **31**:840–854.

Fisher RS (1980) Shotgun wounds. In: Spitz WU and Fisher RS (eds) *Medicolegal Investigation of Death*. Springfield, IL: Charles C Thomas.

Goleb JA and Midkiff CR (1975) Firearms discharge residue sample collection techniques. *Journal of Forensic Sciences* **20**:701–707.

Gunaratnam L and Himberg K (1994) The identification of gunshot residue particles from lead-free Sintox ammunition. *Journal of Forensic Sciences* **39**:532–536.

Harris A (1995) Analysis of primer residue from CCI Blazer® lead free ammunition by scanning electron microscopy/energy dispersive x-ray. *Journal of Forensic Sciences* **40**:27–30.

Jauhari M, Chatterjee SM and Ghosh PK (1972) Statistical treatment of pellet dispersion data for estimating range of firing. *Journal of Forensic Sciences* **17**:141–149.

Krishnan SS (1974) Detection of gunshot residue on the hands by neutron activation and atomic absorption analysis. *Journal of Forensic Sciences* **19**:789–797.

Krishnan SS (1982) Detection of gunshot residue: present status. In: Saferstein R (ed.) *Forensic Science Handbook*, vol 1. Englewood Cliffs: Prentice Hall.

MacCrehan WA, Ricketts KM, Baltzersen RA and Rowe WF (1999) Detecting organic gunpowder residues from handgun use. *Investigative and Forensic Technology* **3576**:116–124.

Mach MH, Pallos A and Jones PF (1978) Feasibility of gunshot residue detection via its organic constituents. Part I: Analysis of smokeless powders by combined gas chromatography-chemical ionization mass spectrometry. *Journal of Forensic Sciences* **23**:433–445.

Mach MH, Pallos A and Jones PF (1978) Feasibility of gunshot residue detection via its organic constituents. Part II: A gas chromatography-mass spectrometry method. *Journal of Forensic Scences* **23**:446–455.

Maiti PC (1973) Powder patterns around bullet holes in bloodstained articles. *Journal of the Forensic Science Society* **13**:197.

Meng HH and Caddy B (1994) Fluorescence detection of ethyl centralite in gunshot residues. *Journal of Forensic Sciences* **39**:1215–1226.

Meng HH and Caddy B (1996) High-performance liquid chromatographic analysis with fluorescence detection of ethyl centralite and 2,4-dinitrotoluene after derivatization with 9-fluorenylmethylchloroformate. *Journal of Forensic Sciences* **41**:215–220.

Moreau TS, Nickels ML, Wray JL, Bottemiller KW and Rowe WF (1985) Pellet patterns fired by sawed-off shotguns. *Journal of Forensic Sciences* **30**:137–149.

Northrop DM and MacCrehan WA (1992) Sample collection, preparation and quantitation in the micellar electrokinetic capillary electrophoresis of gunshot residues. *Journal of Liquid Chromatography* **15**:1041–1063.

Northrop DM and MacCrehan WA (1997) Smokeless powder residue analysis by capillary electrophoresis. *NIJ Technical Report 600-91*. Washington, DC: National Institute of Justice.

Northrop DM, MacCrehan WA and Martire D (1991) Separation and identification of organic gunshot and explosive constituents by micellar electrokinetic capillary electrophoresis. *Analytical Chemistry* **63**:1038–1042.

Reed GE, McGuire PJ and Boehm A (1990) Analysis of gunshot residue tests in 112 suicides. *Journal of Forensic Sciences* **35**:62–68.

Rowe WF and Hanson SR (1985) Range-of-fire estimates from regression analysis applied to the spreads of shotgun pellet patterns: results of a blind study. *Forensic Science International* **28**:239–250.

Rudzitis E (1980) Analysis of the results of gunshot residue detection in case work. *Journal of Forensic Sciences* **25**:839–846.

Schwartz RH and Zona CA (1995) A recovery method for airborne gunshot residue retained in human nasal mucus. *Journal of Forensic Science* **40**:659–661.

Singer RL, Davis D and Houk MM (1996) A survey of gunshot residue analysis methods. *Journal of Forensic Sciences* **41**:195–198.

Spitz WU (1980) Gunshot wounds. In: Spitz WU and Fisher RS (eds) *Medicolegal Investigation of Death*. Springfield, IL: Charles C Thomas.

Stone IC and Petty CS (1974) Examination of gunshot residues. *Journal of Forensic Sciences* **19**:784–788.

Wolten GM, Nesbitt RS, Calloway AR, Loper GL and Jones PF (1977) Final report on particle analysis for gunshot residue detection. *Report ATR-77(7915)-3*. Segundo, California: The Aerospace Corporation.

Wray JL, McNeil JE and Rowe WF (1983) Comparison of methods for estimating range of fire based on the spread of buckshot patterns. *Journal of Forensic Sciences* **28**:846–857.

Weapons, Ammunitions and Penetration

W F Rowe, The George Washington University, Washington, DC, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0535

Introduction

Firearms are heat engines which convert the chemical energy of a propellant into kinetic energy of a projectile weapon. A firearm requires only three components: a projectile, a barrel to confine the hot gases generated by the combustion of the propellant and to guide the projectile toward its target and a source of ignition for the propellant. Since the first appearance of handguns in the fourteenth century a variety of ignition systems have been used: matchlocks, wheel locks, flintlocks and finally percussion locks. Methods of loading firearms have changed from muzzle loading to breech loading. The accuracy of firearms has also been improved by the introduction of rifling, which imparts a stabilizing rotation to the projectile.

Modern Firearms

The variety of modern firearms can be understood by considering the functions that must be performed in order to fire metallic cartridge ammunition. First a round of ammunition must be loaded in the firing

chamber of the weapon. The firing mechanism must be cocked, rendering the weapon ready to fire. Squeezing the trigger actuates the firing mechanism, discharging the round of ammunition. The expended cartridge must be extracted from the firing chamber and then ejected from the firearm so that a fresh round of ammunition may be loaded. In earlier firearms all or most of these functions were carried out manually by the shooter. As firearms designs have progressed more and more of these processes have been rendered automatic with the energy they require being obtained from the firing of the cartridge.

The last decades of the nineteenth century saw a revolution in the development of automatic and semiautomatic (or self-loading) firearms. In these weapons some of the energy released when a cartridge is fired is used to extract and eject the spent cartridge, cock the firing mechanism and load a fresh round of ammunition from a belt or from a magazine. They accomplish this in one of three ways. In the 'locked-breech' system, the weapon's barrel and bolt are locked together and recoil a short distance to the rear as a unit when a cartridge is fired. The barrel is then cammed away from the bolt and the bolt continues to the rear, extracting and ejecting the spent cartridge. The firing mechanism is also cocked by this rearward motion of the bolt. The recoil energy of the bolt is stored (usually in a recoil spring) and is fed back into the bolt. The energy supplied by the recoil spring causes the bolt to move forward, loading a fresh round into the weapon's chamber. Hiram Maxim's machine gun designs (1882–1885) were the first to exploit recoil operation. In 1888 Salvator and Dormus patented the 'blowback' system. In this system the barrel and breechblock are not locked together; the breech is held closed by the pressure of a heavy spring. When a round is fired the inertia of the breechblock and the pressure of the spring hold the breech closed long enough for the bullet to exit the barrel. Then the pressure in the firing chamber forces the breechblock to the rear, compressing the spring. The rearward movement of the breechblock extracts and ejects the cartridge and cocks the firing mechanism. The compressed spring pushes the breechblock forward again, loading a fresh round into the firing chamber. Blowback weapons are chambered for pistol cartridges (e.g. 9 mm parabellum); they cannot fire the powerful cartridges used in rifles. In 1890 John Moses Browning demonstrated the gas-piston system. In gas-operated firearms a small amount of the propellant gases is allowed to escape from a hole in the barrel into a gas piston. The piston is linked to the breechblock by an operating rod; as the propellant gases force the piston to the rear the breechblock is also forced to the rear against a heavy spring. The rearward motion of the breechblock extracts and ejects the spent cartridge and cocks the firing mechanism. The spring forces the breechblock to close and as the breechblock returns to the closed position it loads a fresh cartridge into the firing chamber.

The different types of modern firearms that may be encountered by a firearms examiner are briefly discussed below.

Pistols

These are firearms designed to be held, aimed and fired with one hand (although good pistol marksmanship requires a two-handed grip).

Single-shot and double-barreled pistols Single-shot pistols are usually target pistols. Double-barreled pistols are typically small and easily concealed self-defense weapons.

Revolvers In revolvers, rounds are loaded into firing chambers in a revolving cylinder. The revolver's firing mechanism rotates the cylinder to bring each chamber under the weapon's hammer so that the round in the chamber can be fired. Modern revolvers have between five and eight firing chambers in the cylinder. Loading and unloading is manual, using one of three systems: a loading gate through which each chamber must be individually loaded or unloaded; a swing-out cylinder, whose chambers can be simultaneously emptied using a star-shaped ejector; or a break-top frame which is hinged in front of the cylinder.

Self-loading or semiautomatic pistols Modern semiautomatic pistols operate on either the recoil or blowback system. Magazine capacities vary from six rounds to 33 rounds.

Automatic pistols Some truly automatic pistols have been manufactured, e.g. the 'broomhandled' Mauser (either imitations of the M1912 Mauser produced in Spain or the M1932 Mauser Schnellfeuerpistole), the Beretta M951R, the Glock Model 18 and the Heckler and Koch VP70M. Other pistols designated as 'automatic' are actually self-loading or semiautomatic pistols.

Machine pistols These are actually submachine guns developed in Germany prior to and during World War II, such as the MP38 and the MP40.

Rifles

These are weapons with rifled barrels designed to be fired from the shoulder. Carbines are short-barreled rifles.

Single-shot rifles These are antiques, replicas of antique weapons or target rifles.

Magazine repeaters

Lever action rifles Manipulation of a lever below the breech extracts, ejects, cocks and loads the weapon. This system was developed in the nineteenth century and appeared in the Henry and Winchester repeaters. The system is still popular and a number of sporting rifles use it.

Slide-action or pump-action rifles Pulling a slide to the rear performs extraction, ejection, cocking and loading. This system has been used on some sporting rifles, but has been more popular for sporting shotguns.

Bolt-action rifles There are two types of bolt actions: turn-bolts (which require the lifting of a handle and rotation of the bolt to unlock it from the chamber) and straight-pull bolts (which require the handle be drawn straight to the rear without turning to open the bolt). The Mauser turn-bolt system is the most widely used bolt action.

Self-loading or semiautomatic rifles The M1 Garand military rifle, which operates on the gas-piston principle, was the first semiautomatic rifle to be adopted as the standard infantry weapon by a major military power. There are currently a large number of military and sporting semiautomatic rifles available.

Automatic rifles Automatic rifles were developed during World War I to provide advancing infantry close supporting automatic fire, contemporary machine guns being too large and heavy to be moved forward rapidly. The Browning automatic rifle (BAR) is an example of this type of weapon. Because automatic rifles were heavy and had magazines with limited capacities they were eventually replaced by light machine guns and assault rifles.

Assault rifles These are light automatic weapons which fire shortened or reduced charge rifle cartridges from magazines. Many have selectors which permit switching between automatic and semiautomatic modes of fire. The first assault rifles (or *Sturmgewehr*) were developed to give advancing infantry greater firepower. Because they fire rifle cartridges their rounds have greater striking power then submachine guns. Examples of assault rifles include the German MP43 (the original *Sturmgewehr*) and the Kalashnikov AK-47.

Submachine guns

These are lightweight automatic weapons; they fire pistol ammunition from magazines. Semiautomatic versions of some submachine guns such as the UZI and the Sterling are also marketed.

Machine guns

These are automatic weapons in which the ammunition is fed from a magazine or a belt. Because machine guns fire more powerful cartridges than submachine guns or assault rifles, they are usually fired from a support such as a bipod or tripod or are mounted in a vehicle such as a tank, an armored personnel carrier or an airplane. Infantry machine guns are usually designed to permit rapid changes of barrel (to avoid overheating and excessive wear of the barrel during prolonged firing); machine gun squads usually carry spare barrels and frequently change them during firefights. A number of modern armies use machine gun designs based on the German MG42 general-purpose machine gun; the US M60 general purpose machine gun combines features of the MG42 and the German FG42 automatic rifle.

Shotguns

These are weapons with smoothbore barrels designed to be fired from the shoulder. Their ammunition typically contains small pellets.

Single-shot shotguns This type of shotgun is not as popular in the United States as double-barrel, slide-action or semiautomatic shotguns; they are however quite common in Europe and the Third World.

Double-barreled shotguns The two barrels are either side by side or one above the other (over–under shotguns). Double-barrel shotguns usually break open at the top for loading and reloading.

Magazine repeaters

Lever-action shotgun This is an uncommon type of shotgun.

Slide-action or pump action shotgun This is one of the more popular types of shotgun.

Bolt action-action shotgun This is an uncommon type of shotgun.

Self-loading or semiautomatic shotgun This type of shotgun has become one of the most popular types.

Improvised weapons

Firearms can be improvised from a variety of components. 'Zip gun' is a term applied to firearms improvised from such unpromising material as automobile aerials (for barrels), wooden blocks (for grips), nails (for firing pins) and rubber bands (for main springs). Childrens cap pistols have been modified to fire metallic cartridges as have blank starter pistols. Pocket teargas pins have also been modified to fire cartridges.

Manufacturing Methods

Because firearms examiners must interpret the markings made on fired bullets and cartridges by various parts of a firearm these forensic scientists should have an understanding of the methods used to manufacture firearms and ammunition. Bullets are marked by the rifling in the weapon's barrel. Conventional rifling consists of lands and grooves that impart a stabilizing rotation to the fired bullet. There are five processes that have been used to rifle gun barrels.

There are two cut-rifling methods. The *hook and scrape cutters* are the oldest methods for rifling barrels and may still be used by hobbyists. The barrel is prepared for rifling by drilling a hole through the center of a piece of steel bar stock. A hook cutter or scrape cutter is then used to cut the grooves to the desired depth. The hook cutter is an oblong piece of metal with a hook-shaped cutting edge. A metal rod is used to draw the hook cutter through the barrel, shaving away the metal in a groove until the desired depth is reached. As the hook cutter is drawn through the barrel is rotated so that the groove spirals down the barrel. The hook cutter cuts one groove at a time. The scrape cutter functions in a similar fashion; however, the cutting surface of the scrape cutter is a raised ridge. The scrape cutter can have a cutting surface on both sides so that two grooves on opposite sides of the barrel can be cut simultaneously.

The *gang broach* consists of a series of cutting tools mounted on a metal shaft. Each tool has cutting surfaces corresponding to each of the grooves. The barrel is prepared for rifling by drilling out a piece of steel bar stock. The gang broach is then forced by hydraulic pressure through the barrel, each cutting tool in succession cutting the grooves a little deeper until the desired groove depth is reached. All the grooves in the rifling are cut with a single pass of the gang broach.

Three other rifling methods are also used. In *swaging* a rifling button composed of tungsten carbide is forced down the drilled-out barrel, simultaneously engraving the lands and grooves of the rifling on the inside of the barrel and expanding the bore to its finished diameter. The metal on the interior surface of the barrel is hardened by cold-working. Inexpensive smallbore firearms are produced by swaging.

In *hammer forging*, a piece of steel bar stock is drilled out to a diameter greater than the intended finished bore diameter. The drilled stock is then slipped over a hardened metal mandrel and hammered down. The mandrel is then forced out of the rifled barrel. Hammer forging is used to manufacture the barrels of Glock pistols. These pistols do not have conventional land-and-groove rifling; instead the interiors of their barrels have hexagonal cross-sections.

During the process of *electrochemical etching*, resin layers are painted on to the interior of the drilled out barrel, corresponding to the desired configuration of lands. A chemical solution is then poured into the barrel and used to etch the grooves electrochemically into the interior of the barrel.

Cut-rifling methods produce barrels that have unique microscopic imperfections that produce unique striation patterns on fired bullets – even when the weapon is new. In the case of hook and scrape cutting and broaching the rifling tools wear just enough in rifling a barrel that barrels rifled in succession with the same rifling tools leave distinguishable markings on fired bullets. Swaging and electrochemical etching also appear to produce unique microscopic features within the gun bore. Use of the firearm creates new imperfections in the barrel that mark the fired bullets. These are caused by erosion and corrosion of the gun bore. Erosion refers to the wearing away of the metal of the barrel by friction; corrosion refers to the chemical attack on the metal comprising the barrel by propellant residues. The gradual change in the surface imperfections of the gun barrel can result in a failure to match a questioned bullet to the weapon that fired it, particularly if the weapon has been extensively used between the time the questioned bullet was fired in the commission of the crime and the time it was seized by police as evidence.

Other firearm components of particular interest to firearms examiners are the firing pin, the breechblock or backing plate, the extractor and the ejector. The firing pin strikes the percussion cap of centerfire cartridges or the rim of rimfire cartridges in order to detonate the primer compound in the cartridge. Firing pins are finished by hand-filing or by turning on a lathe (a process that also involves hand-filing). These handfinishing operations impart unique microscopic features to the firing pin surface; moreover, firing pins are prone to acquire cracks and other damage from use and abuse. Firearms other than

revolvers have a breechblock against which the base of the cartridge rests prior to firing. Upon firing the cartridge is forced back against the breechblock; the primer cap in centerfire cartridges and sometimes even the cartridge casing itself will be marked by the surface of the breechblock. Breechblocks are handfiled, milled, turned on a lathe, cut with a broach or stamped. Most of these processes produce a unique pattern of marks which is often referred to as the boltface signature. In the case of revolvers the portion of the frame immediately behind the top chamber in the cylinder (called the backing plate) may similarly mark fired cartridges.

Firing pin marks and breechblock/backing plate markings generally have the greatest significance for firearms examiners because they can only be produced by firing a cartridge in the weapon. Other markings such as extractor and ejector markings may be produced when a cartridge is merely loaded into a weapon and then extracted and ejected manually. In automatic and semiautomatic weapons the extractor is a hook that engages the base of the cartridge and pulls the expended cartridge from the weapon's firing chamber. Extractors are necessary because firing causes cartridges to expand in the chamber and bind against its walls. Ejectors are usually projections in automatic or semiautomatic firearms that wrest the fired cartridge from the extractor so that it is ejected from the weapon. The term ejector is also used to refer to the devices that are used to remove fired cartridges from the cylinders of revolvers. They take the form either of a rod that pushes the cartridges one at a time from the chamber in the cylinder or a star-shaped piece of flat metal that is used to remove all the cartridges from the cylinder simultaneously. In general, revolver ejectors do not produce forensically useful markings. The ejectors and extractors in semiautomatic and automatic weapons are finished by handfiling.

The Class Characteristics of Firearms

The class characteristics of firearms are those features or attributes that identify the make and model. One task of the firearms examiner is to determine sufficient class characteristics from fired bullets or cartridges to identify the make and model of firearm that fired them. This is not always possible. Most of the time the firearms examiner must be satisfied to limit the makes and models of firearms that could have fired a bullet or cartridge to a small number. The class characteristics determinable from bullets and cartridges are listed below.

General rifling characteristics

Caliber The caliber of a firearm is nominally the bore diameter, i.e. the diameter of a circle that just touches the tops of the lands of the rifling. This diameter may be expressed in hundredths or thousandths of an inch (e.g. 0.38 caliber or 0.380 caliber) or in millimeters (e.g. 5.56 mm or 7.62 mm). It is important to keep in mind that these bore diameters are merely nominal: a weapon designated as 0.38 caliber may have an actual bore diameter that is much smaller. For example, Smith and Wesson revolvers manufactured in a nominal 0.38 caliber have bore diameters ranging from 0.3463 inches to 0.3515 inches. Colt semiautomatic pistols manufactured in a nominal 0.45 caliber have bore diameters ranging from 0.4422 inches to 0.4461 inches. The groove diameter (i.e. the diameter of a circle that just touches the bottom of the grooves) can also be measured and many tabulations of firearm rifling characteristics include it along with the bore diameter.

Number of lands and grooves The number of lands and grooves in the barrel of a rifled firearm can range from as few as two to more than twenty.

Direction of twist of the rifling The rifling in a firearm barrel can spiral in one of two directions. For right twist (or Smith and Wesson) type rifling the thumb of the right hand points in the direction the bullet travels, while the other fingers curl in the direction of the rotation of the bullet. For left twist (or Colt) type rifling the thumb of the left hand points in the direction the bullet travels, while the other fingers curl in the direction of rotation of the bullet.

Degree of twist of the rifling This is the angle made by the lands of the rifling with the axis of the barrel. It is often expressed as the number of inches of barrel required for one complete turn of the rifling.

Class characteristics of cartridges

Shape of firing chamber Cartridges may have straight sides or may be bottle-shaped to fit the firing chamber. Some firing chambers have flutes which are used to 'float' the expended cartridge out of the firing chamber; flutes leave characteristic markings on the sides of fired cartridges.

Location of firing pin Modern firearm cartridges are either rimfire or centerfire.

Size and shape of firing pin For 0.22 caliber single-shot 'boys rifles' the size and shape of the firing pin is one of the most diagnostic class characteristics. Some

other weapons (Beretta 1934, Glock pistols and the Bren gun) produce very distinctive firing pin impressions. For other types of weapon this class characteristic is less valuable.

Size and shape of extractors and ejectors Among automatic and semiautomatic firearms there is considerable variation in the size and shape of the extractor and ejector. For example, the 0.45 caliber M1911A1 semiautomatic pistol has a narrow strip of spring steel for an extractor, whereas the M14 semiautomatic rifle has two broad extractors on either side of the bolt.

Geometrical relationship of extractor and ejector This determines the direction in which expended cartridges are ejected from the weapon. Extractors and ejectors may be 180°, 120° or 90° apart.

Shotguns

The barrels of shotguns lack rifling; they are usually produced from drilled out bar stock or from extruded seamless metal tubing. The diameter of a shotgun barrel is determined by its gauge. The gauge of a barrel is the number of lead balls whose diameters are the same as the interior diameter of the barrel whose total weight is one pound. A 10 gauge shotgun has a bore diameter of 0.775 inch; a 12 gauge shotgun barrel has a diameter of 0.730 inches. In the case of 410 gauge the gauge is the actual bore diameter. Because shotgun barrels lack rifling shotguns will not normally produce markings on the projectiles that they fire. Criminals may saw off the barrel of a shotgun to facilitate concealment. Burrs may be left which will mark the sides of shot columns or the edges of wads. An important feature of shotgun barrels is the presence of a choke or constriction at the muzzle. The purpose of the choke is to concentrate the shot pattern so that enough shotgun pellets strike the game animal to inflict a fatal or incapacitating wound. The degree of constriction is indicated by the designation of the barrel as cylinder bored (no choke), improved cylinder (slight constriction), modified choke (some constriction) and full choke (maximum constriction). The degree of choke of a shotgun barrel is determined by the fraction of pellets falling within a 30-inch (76.2 cm) circle at a range of 40 yards (36.6 m). A cylinder-bore shotgun will place 25–35% of the pellets inside the circle; an improved cylinder barrel 35–45%; a modified choke barrel 45–65%; and a full-choke barrel 65–75%. The barrels of double-barreled shotguns usually have different chokes. The choke of a shotgun barrel may be increased by the use of barrel inserts or adjustable compensators (which also direct the propellant gases issuing from the muzzle to the side and rear to reduce recoil). The choke of a shotgun affects the size of the pellet pattern it fires; consequently the choke of a shotgun will have an effect on the determination of the range at which a questioned pellet pattern was fired.

Ammunition

Types of bullets

Bullets are made in a variety of shapes (e.g. round nosed, pointed nose and wad cutter) and of a variety of materials. Lead bullets are made of relatively pure lead; because of their softness they are used in low-velocity firearms (0.22 caliber pistols and rifles). Lead alloy bullets contain a few percent of an alloying element such as antimony (commercially manufactured bullets) or tin (home-cast bullets). Lead alloy bullets are harder than lead bullets and are used in higher muzzle velocity firearms. Both lead and lead alloy bullets readily deform on impact; they may also fragment. Severe deformation and fragmentation can severely handicap the firearms examiner. Lead and lead alloy bullets are also subject to oxidation that can obscure or obliterate surface striations. On the positive side, lead and lead alloy bullets expand to completely fill the rifling of the weapon so that the firearms examiner may have a very large number of striations which can be matched.

Full-metal jacket bullets consist of a lead core covered by a thin sheet of gilding metal, brass, steel or aluminum. Usually the lead core is exposed at the base of the bullet. Full-metal jacket bullets resist deformation and fragmentation better than lead or lead alloy bullets. On the other hand, however, full-metal jacket bullets do not expand to completely fill the weapon's rifling with the result that such bullets are not as well-marked as lead or lead alloy bullets. Luballoy® bullets consist of a lead core covered by a very thin layer of copper. The copper layer tends to flake off the bullets' surfaces, taking with it the fine striation patterns.

Semijacketed or partially jacketed bullets also have lead cores; commonly, their sides and bases are covered with a metal jacket, leaving the lead core exposed at the nose. There is a variety of semijacketed bullets, including the following.

1. Semijacketed hollow-point bullets in which the nose of the bullet is hollowed out to facilitate expansion of the bullet. Exploding bullets are

hollow points with a percussion cap inserted into the cavity in the nose. Metal spheres may also be inserted in the cavity in the nose to facilitate expansion.
2. Soft point bullets have a soft metal plug placed in the nose of the bullet to facilitate expansion on impact.
3. Metal point bullets have a hardened metal cap on their noses; their sides are not jacketed.
4. Bronze point bullets, in order to facilitate bullet penetration and expansion in thick-skinned game animals, have a pointed bronze casting inserted in the nose of the bullet. The sides of the bullet are normally jacketed.

In addition to the bullet types discussed above there are a number of specialty bullets. Frangible bullets are composed of particles of powdered iron pressed together or particles of powdered iron and lead pressed together with an organic binder. These bullets break up on impact so that they present no danger of ricochet. Frangible bullets are used in shooting galleries and for cattle stunning in slaughterhouses. Glaser safety slugs were developed to deal with the hazards associated with ricochet and 'overpenetration' of conventional bullets. Conventional bullets which miss their intended human target may ricochet and strike an unintended target; conventional bullets may go through the intended target and strike a second target. The projectile in Glaser safety slugs is a plastic bag filled with shot. When it impacts a target the bag spreads out, distributing the impact forces. At close range the Glaser safety slugs can incapacitate their human target; however, they are unlikely to pass through the body and injure a second person. Armor-piercing bullets have steel jackets which are poorly marked by the rifling of the gun barrel; these bullets may have tungsten carbide cores. Teflon-coated bullets are covered with a layer of self-lubricating fluorocarbon polymer. This layer reduces friction between the bullet surface and the interior of the gun barrel so that these bullets are able to reach very high muzzle velocities. Teflon-coated bullets were designed to aid police officers in situations where criminals have taken shelter behind an automobile; these bullets can penetrate an automotive engine block and can also penetrate soft body armor. Open tubular bullets can also achieve very high muzzle velocities. These bullets consist of an open tube of copper or brass with a plastic cup which acts as a gas check. Open tubular bullets made of brass can penetrate soft body armor.

Discarding sabot rounds have a semijacketed soft-point bullet pressed into a plastic cup or sabot. The Remington Accelerator cartridge is a 0.30/06 rifle cartridge with a 5.56 mm (0.223 caliber) semijacketed soft-point bullet. When this type of round is fired the sabot separates from the bullet and eventually falls to the ground. Only the sabot engages the rifling in the weapon's barrel; therefore, it alone bears class and individual characteristics.

Cartridge cases

Cartridge cases are made of brass, steel or aluminum. They may be plated with nickel to reduce corrosion. Cartridges are made in a variety of shapes (e.g. straight-sided or bottle-shaped) to conform to the different shapes of the firing chambers of firearms. There are a number of broad categories of cartridge shapes.

1. Belted cartridges are high-powered rifle cartridges. They have a raised belt at the base of the cartridge to control the positioning of the cartridge in the firing chamber. Belted cartridges may be rimmed or rimless (see below).
2. Rimmed cartridges have a rim at the base. Rimmed cartridges are used in revolvers; the rims prevent the cartridges from sliding forward out of the firing chambers of the revolver's cylinder.
3. Rimless cartridges have a base that is the same diameter as that of the body of the cartridge. Rimless cartridges have an *extractor groove* circling the base of the cartridge. The extractor of the firing weapon engages this groove so that the expended cartridge can be extracted from the weapon.
4. Semirimmed cartridges have an extractor groove like that found in rimless cartridges; the base of the cartridge has a larger diameter than the body of the cartridge.
5. Rebated (or rebated-rimless) cartridges resemble rimless cartridges in that they have an extractor groove; however, the base of the cartridge has a diameter smaller than the body of the cartridge.

Cartridges can also be categorized by the location of the primer that is used to initiate the propellant in the cartridge. In rimfire cartridges the primer compound is placed inside the rolled rim of the cartridge; the weapon's firing pin strikes the rim of the cartridge to fire it. Centerfire cartridges have a depression in the center of the cartridge for a percussion cap made of gilding metal. *Boxer* cartridges are used in the United States; these have a cylindrical depression in the center of the base of the cartridge for the percussion cap with a flashhole leading to the powder. The Boxer primer contains a small metal anvil against which the firing pin crushes the primer compound. *Berdan* cartridges are used in Europe and other parts of the world. The Berdan cartridge has a cylindrical depression in its

base. There is a small cone in the center of the depression through which three flashholes lead to the powder. The cone fulfills the role of the anvil in the Boxer primer; the firing pin crushes the primer compound against the cone. The base of the cartridge bears a headstamp. The headstamp indicates the vendor of the cartridge and the type of cartridge. The headstamp on a rifle cartridge might read 'REM-UMC 30-06 SPRG' which would reveal that the cartridge was manufactured by Remington–United Metallic Cartridge Company and that it is a bottle-shaped cartridge introduced in 1906 for 0.30 caliber Springfield rifles.

Shotshells

The bodies of shotshells are made of coated paper or plastic. They have bases made of brass. Shotshells have headstamps like the cartridges used in other types of firearms. Shotgun ammunition normally contains a mass of small pellets. The pellets may be lead, lead alloy ('chilled' shot) or stainless steel. Their diameters vary from 0.36 inches (000 buckshot) down to 0.05 inches (No. 12 birdshot). The number of pellets per shotshell is related to their size and the gauge of the shotgun. For example, 12 gauge 00 buckshot shotshells contain nine or twelve 00 buckshot, whereas 12 gauge no. 2 shotshells may contain over a hundred no. 2 birdshot. Wads are disks of cardboard, thin wood or plastic that are used to hold the shot in place inside the shotshell. One wad will typically separate the shot from the propellant (the overpowder wad); another wad may be placed on top of the shot mass at the mouth of the shotshell (the overshot wad). In many shotshells the wads may be replaced by plastic one- or two-piece shot columns. These shot columns have plastic cups that enclose the shot and protect them from deformation due to contact with the interior of the shotgun barrel. Plastic particles may be packed with the shot to provide similar cushioning. Wads and shot columns may be found inside shotgun wounds inflicted at close range. Patterned contusions caused by wads and shot columns may occasionally be seen adjacent to shotgun wounds inflicted at ranges of several feet.

Shotshells may also contain single projectiles such as round balls or rifled slugs. The single round balls fired by shotguns are essentially the same as the round balls once fired in smoothbore muskets. Rifled slugs are hollow-base conical lead slugs with projecting fins. The fins provide stabilization for the slugs so that they are more accurate than round balls. The wounding effects of rifled shotgun slugs are similar to those produced by the Minie bullets (or minnie balls) used during the American Civil War.

Powders

The original propellant used in firearms was black powder (the name being a translation of Schwartzpulver, a propellant formulation named for Berthold Schwartz), originally a mixture of charcoal, sulfur and potassium nitrate. Black powder was used as the propellent in firearms until the end of the nineteenth century when it was gradually replaced by smokeless powder. Hobbyists (black powder enthusiasts, Civil War re-enactors and the like) still use black powder or black powder substitutes (e.g. sulfurless black powder or Pyrodex) as propellants. Smokeless powders contain nitrocellulose. Single-base smokeless powders contain nitrocellulose combined with various additives (e.g. potassium sulfate, diphenylamine and dinitrotoluene) which affect the mechanical or chemical properties of the powder grains. Double-base smokeless powders contain nitroglycerin as an additional energetic material. The grains of smokeless powders are produced either by the extrusion process or by the ball powder process. In the extrusion process, the ingredients of the propellant are mixed together in a cake with a volatile solvent. The cake is forced against a steel plate in which a number of holes have been made. The propellant mixture is extruded through these holes and a rotating metal blade cuts off short lengths of extruded material. Extruded propellant grains may take a variety of forms: rods, perforated rods, flattened disks, perforated disks or flakes. The ball powder process also involves an extrusion process, except that the propellant grains are extruded into hot water where they assume spherical shapes. The spheres may be further processed by flattening them between metal rollers. Propellants produced by the ball powder process are either spherical or flattened disks.

Propellants for firearms are manufactured in batches. The manufacturer tests each batch to determine its ballistic performance. To achieve the ballistic performance desired for the final product two or more batches of powder will be blended together. Thus, the powder in a cartridge or shotshell will consist of grains from different batches having slightly different chemical compositions.

See also: **Firearms**: Range; Residues; Humane Killing Tools; Laboratory Analysis. **Pattern Evidence**: Shotgun Ammunition on a Target.

Further Reading

Berg SO (1977) The forensic ballistics laboratory. In: Tedeschi CG, Eckert WG and Tedeschi LG (eds) *Forensic Medicine: A Study in Trauma and Environmental Hazard*, vol. 1. Philadelphia: W. B. Saunders.

Ceccaldi PF (1962) Examination of firearms and ammunition. In: Lundquist F (ed.) *Methods of Forensic Science*, vol. 1. New York: Interscience.

Davis JE (1958) *An Introduction to Tool Marks, Firearms and the Striagraph*. Springfield: Charles C Thomas.

Ezell EC (1983) *Small Arms of the World: A Basic Manual of Small Arms*, 12th edn. Harrisburg: Stackpole Books.

Ezell EC (1986) *The AK47 Story: Evolution of the Kalashnikov Weapons*. Harrisburg: Stackpole Books.

Ezell EC (1988) *Small Arms Today: Latest Reports on the World's Weapons and Ammunition*, 2nd edn. Harrisburg: Stackpole Books.

Hatcher JS, Jury FJ and Weller J (1957) *Firearms Investigation, Identification and Evidence*. Harrisburg: Stackpole Books.

Hogg IV (1977) *The Encyclopedia of Infantry Weapons of World War II*. New York: Thomas Y. Crowell.

Hogg I (1996) *The Story of the Gun*. New York: St Martin's Press.

Hogg IV (1999) *The Greenhill Military Small Arms Data Book*. Mechanicsburg: Stackpole.

Hogg I and Adam R (1996) *Jane's Guns Recognition Guide*. Glasgow: HarperCollins.

Jinks RG (1991) *Artistry in Arms: the Guns of Smith & Wessson*. Springfield: Smith and Wesson.

Kettrick JF (1983) *Small Arms Identification and Operations Guide – Eurasian Communist Countries (U)*. Washington: Defense Intelligence Agency.

Kettrick JF (1984) *Small Arms Identification and Operations Guide – Free World*. Washington: Defense Intelligence Agency.

Koller L (1953) *Complete Book of Guns*. New York: Maco Magazine Corporation.

Krcma V (1996) Fluted and annular grooved barrel chambers in firearms. *Journal of Forensic Sciences* **41**:407–417.

Mathews JH (1962) *Firearms Identification*, vols 1–3. Springfield: Charles C Thomas.

Nichols RG (1997) Firearm and toolmark identification criteria: a review of the literature. *Journal of Forensic Sciences* **42**:466–474.

Rowe WF (1988) Firearms identification. In: Saferstein R (ed.) *Forensic Science Handbook*, vol II. Englewood Cliffs: Prentice Hall.

Smith WHB (1997) *Pistols and Revolvers*. Birmingham: Palladin Press.

Smith WHB (1997) *Rifles*. Birmingham: Palladin Press.

Footprints *see* **Pattern Evidence**: Bare Footprint Marks.

Footwear *see* **Pattern Evidence**: Footwear Marks.

Forensic Science *see* **Criminalistics**.

FORGERY AND FRAUD

Contents

Counterfeit Currency
Payment Cards

Counterfeit Currency

M J Surrency, **L W Pagano** and **S E Church**, US Secret Service, Washington, DC, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0547

Introduction

Since ancient times and across civilizations, many types of items, including stones, shells, spices, chocolate, beads, livestock and various metals, have been used as symbols of value for purposes of trade. It was not until after the seventeenth century that printed paper came into use as the chief instrument for direct transactions. When Marco Polo brought back tales that paper currency served as money in China in the thirteenth century, he was met with disbelief in the West, where trade was based on coins, crops and service. Yet, by this time, paper money in China already dated several centuries to the T'ang dynasty. The use of paper currency survived in China until the end of the fourteenth century (under the Ming dynas-

ties), after which its use was abandoned. In China, the long-standing use of paper money was linked to the heavy-handedness of the centralized government in forcing its acceptance and in refusing alternative sources of personal value.

In contrast, in the West, the first appearances of paper currency were prompted by emergency situations in the form of coin shortages. In Sweden in 1661, a shortage of silver forced the first appearance of paper money but its use was met with distrust and fear that the nation's now overly portable wealth would quickly flee the country in the pockets of foreigners. A few years later, paper money was introduced in North America when a delay in the shipment of soldiers' pay in Canada prompted authorities to monetize and distribute playing cards – the highest quality paper available – overprinted with monetary values, official seals and signatures in lieu of the normal payments in the coin of the realm.

Counterfeiting

As long as money has existed, so has counterfeiting. In fact, although the US Constitution does not specifically authorize the federal government to issue currency, it does specifically authorize punishment for counterfeiting (Article 1, section 8). Penalties for counterfeiting have often been severe, to include execution. Some banknotes have even carried printed warnings such as 'To Counterfeit Is Death'. In fact, counterfeiters are still executed in China today. In the USA, however, the maximum sentence is not so harsh: 15 years imprisonment and $5000 fine.

There were two periods in US history when floods of counterfeit overwhelmed the paper currency supply (during the Continental period and the decades before the Civil War) and threatened its demise. In the middle part of the nineteenth century, so-called 'counterfeit detectors' were commonly used. These were handbooks that described all the genuine notes in circulation, which were issued by state and local banks, and explicated counterfeit flaws in great detail. By the start of the Civil War, an estimated one-third of the currency in circulation was fraudulent.

This counterfeiting problem, along with the need to finance the Civil War, led the US Government in 1863 to authorize a national currency, to be manufactured by the US Treasury Department's newly established Bureau of Engraving and Printing. By 1865, the need for an enforcement agency to fight counterfeiting and protect the integrity of the new currency led to the establishment of the US Secret Service.

The US Secret Service has been extremely effective in carrying out its responsibilities. Over the course of the twentieth century, the US dollar became so stable and trustworthy that it came into popular use throughout the world as not only an instrument of trade, but also of stored value. Thus today, due to the stability of the currency, the economy and the society, the familiar US 'greenback' serves as a *de facto* world currency. In fact, at the end of the 20th century almost $600 billion in paper currency is in circulation worldwide, and an estimated 60% is held abroad, escalating the importance of protecting US currency from counterfeiting.

Genuine Currency

The materials and characteristics of genuine banknotes are produced to certain specifications, but no such standards exist for counterfeit banknotes. Individual counterfeiters set their own standards, and these standards vary widely. Because the counterfeits lack standards, the process of their identification becomes one of proving a negative, i.e. that the specimen is not genuine. Thus, knowledge of the genuine specimen is critical to identifying counterfeits and evaluation is best carried out by direct, side-by-side comparison of the suspect document with a genuine specimen.

Security

Genuine banknotes are designed to be difficult to counterfeit. They are made with special materials applied in ways not easily reproduced, or accessible, to the general public. Banknotes have traditionally been printed by intaglio, offset and typographic printing processes on high-quality rag paper, using highly durable inks in an integrated design with a redundancy of information, especially regarding the value. The design and materials are chosen to be complementary, each reinforcing the other, and to be challenging for counterfeiters to reproduce.

Durability

The specifications for genuine banknotes are determined not only by the need for resistance to counterfeiting, but also for resistance to wear. In its lifetime, a banknote will be handled daily, in multiple transactions, so its construction and materials must be highly durable and resistant to wear. This requirement extends not only to the base substrate and ink, but also to any special design elements, including applied features such as security threads and fibers, reflective or holographic foils, optically active materials and machine-detectable features.

Production Method

Paper/substrate

The traditional substrate for security documents is a highly durable, rough, uncoated rag paper made with cotton, linen or a blend of cotton and linen; other pulps and materials may be added to enhance strength or tear properties, to reduce costs or to increase security. These additional materials may be fillers and binders, other cellulosic fibers to supplement the base materials, or distinctive materials such as natural or synthetic security fibers or planchettes (distinctive tissue or polymer disks added to the paper furnish) to distinguish the security paper from generic rag stock. The security fibers or planchettes may include special materials for overt, covert or forensic identification. The color of the paper may be white, off-white or other pastel shades that complement the design. A typical basic weight of the paper is 80–90 g m^{-2}, although some banknote papers weigh in much lower and some higher. By tradition, security papers do not fluoresce under ultraviolet (UV) light stimulation, but the inclusions (security fibers, security threads and planchettes) and the inks may. Security threads, which are nonpaper materials (plastic, metal or actual threads), are inserted during the papermaking process and can be either entirely within the paper sheet or rise occasionally to the surface for a 'windowed' effect. Security threads are often used as convenient vehicles for introducing machine-readable features or more overt characteristics such as legible text or fluorescence properties.

Security papers often include large portrait watermarks registered in the sheet to fall consistently in a localized area within the printed image. In some currencies, the watermark represents an animal or other easily recognizable object that fits with the theme of the note design. These designs are formed directly in the papermaking process, in which areas of greater and lesser fiber density are created to give darker and lighter areas of the paper.

Today, alternate substrates made from polymers or polymer–paper composites are in use or being introduced.

Printing processes

The three printing processes typically used in printing currency are intaglio, typographic (or letterpress) and offset (both wet and dry). Typically at least the main image (often a portrait) is rendered using intaglio, with the vignettes, background details, see-through images and often the reverse sides printed by offset; the serial numbers are added typographically. In some currencies, the images are printed using only offset; US currency is printed using only intaglio and typographic printing.

Intaglio Traditionally, intaglio printing is used mainly for engraved artwork, special announcement cards, invitations and banknotes. In this method, the image to be printed is engraved as lines into a metal plate. Special intaglio ink, which is very viscous and opaque, is forced into the engravings, but wiped from the plate surface. It is then transferred under high pressure to the intended substrate.

The utility of intaglio printing for security documents lies in the characteristic raised, embossed and feathered texture of the image that is not present when other methods are used. Microscopically, the ink can be seen to actually sit on the surface of the paper and spread along individual paper fibers (feathering), giving the image both visual and tactual texture (**Fig. 1**). These textural characteristics are especially important because anecdotal evidence suggests that counterfeits, whether on paper or polymer substrates, are often first questioned because they 'just don't feel right'.

Typographic (letterpress) In the typographic, or letterpress, printing process, ink is applied to a raised surface and transferred by direct impression to the substrate. This method of ink transfer creates an effect known as 'squeeze out', in which the excess ink on the printing surface is forced toward the edges of the image on to the paper, creating a thick ink outline of the image (**Fig. 2**). These characteristics are dependent on the type of plate used, the printing pressure, and the substrate on which the image was printed.

Typical examples of this process are typewriters and date stamps. The utility in banknote printing is that, since the image to be printed can be changed for

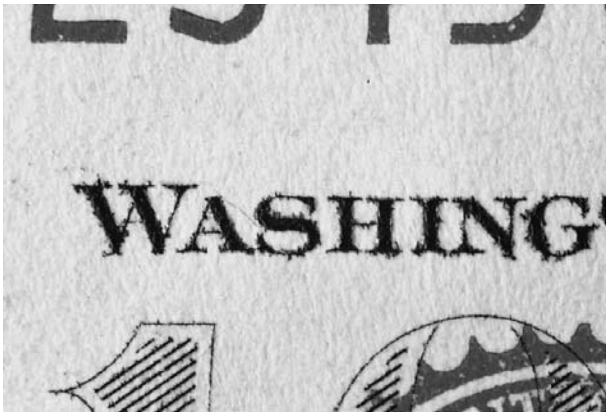


Figure 1 Intaglio printing on genuine US banknote.

each impression, typographic printing can be used to add unique identifiers such as serial numbers or bar codes.

Offset Offset lithographic printing is a planographic process, with neither raised nor engraved features in the printing plates. It is widely used in the commercial arena, especially for printing newspapers and other high-volume products. The image is created in a single surface, with the coplanar image and nonimage areas distinguished chemically so that the image parts will accept ink, but the nonimage areas will not. The ink is transferred from the treated plate on to an intermediate transfer (offset) blanket, then to the target substrate.

Microscopically, the ink appears sharp and even, without build-up of ink near the edges or middle. Since the process is planographic, the printed paper lacks any distortions, such as the indentations or embossing characteristic of letterpress or feathering of intaglio (**Fig. 3**).

Although a commonly used process, offset printing fills a special niche in security document printing. The modern, so-called perfecting presses used by banknote printers are capable of extremely close color-to-color and front-to-back registration. With this equipment, specially designed icons can be printed half on one side of the note and half on the other, forming the complete image only in transmitted light, when both halves can be seen. In addition, these presses are used to apply very fine and interwoven lines in a range of color saturations that challenge the traditional print-shop counterfeiter and modern reprographic equipment as well. With offset printing, the range of microprinted features is extended beyond the capabilities of intaglio.

Another offset process used by banknote printers is dry offset. Not to be confused with waterless offset, where the planographic image and nonimage areas are separated chemically, dry offset uses a raised image plate in combination with an intermediate transfer blanket. The resulting print often exhibits characteristics of letterpress, but lacks any embossment of the paper.

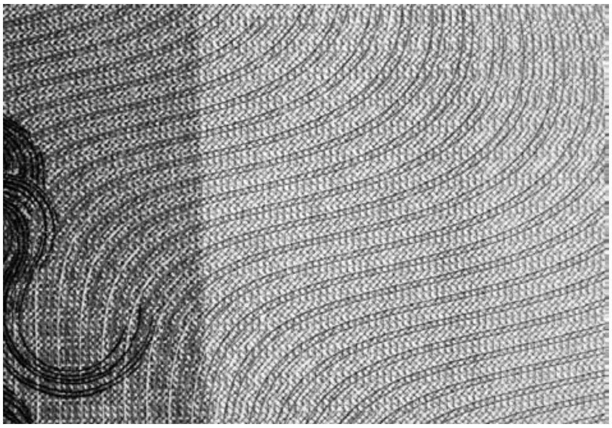


Figure 3 Offset printing.

Screen Screen printing utilizes a stencil to separate the image and nonimage areas. The stencil is fixed to a screen support and allows ink to pass through the image areas to the substrate/paper. The screen pattern is often apparent in the images produced by this method (**Fig. 4**). The ink used for screen printing is thick and the image raised, similar to intaglio, but since no significant pressure is applied to the paper during printing, it lacks any embossing.

Inks

The inks used in security printing are chosen not only for their colour and printing properties, but also for their wear characteristics. Durability is important

Figure 2 Typographic printing on genuine US banknote.

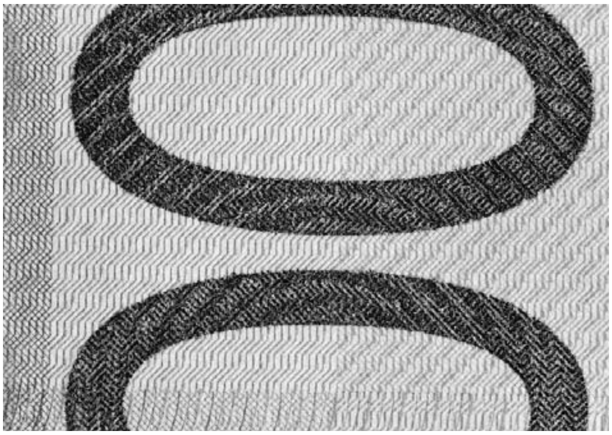


Figure 4 (see color plate 26) Screen printing.

from two aspects. From a utilitarian standpoint, banknotes must withstand the repeated use and occasionally harsh treatment, even laundry and burial, received at the hands of the public. Therefore, to maintain its integrity and utility, a banknote must be sturdy. From a security aspect, the easy removal of the inks from a banknote leaves it vulnerable to counterfeiting abuse, enabling easier color-by-color reproduction or overprinting of higher denominations. For these reasons, inks used in printing banknotes are subjected to a full range of durability tests, including crumpling, laundering and chemical resistance.

Security inks are also flexible and convenient carriers for optical, magnetic or other machine authentication features. Recently, color-shifting inks have come into use. These inks, which use optically variable flakes as pigments, appear to be one color normal to the printed surface, but another color at an oblique angle (**Fig. 5**). Other inks may have special properties such as visible emission under ultraviolet illumination, or certain spectral properties that lead to different appearances under different lighting conditions, or other more covert features for highly secure authentication. The inks used by banknote printers are usually specially designed for the purpose and are not generally available outside the security printing industry.

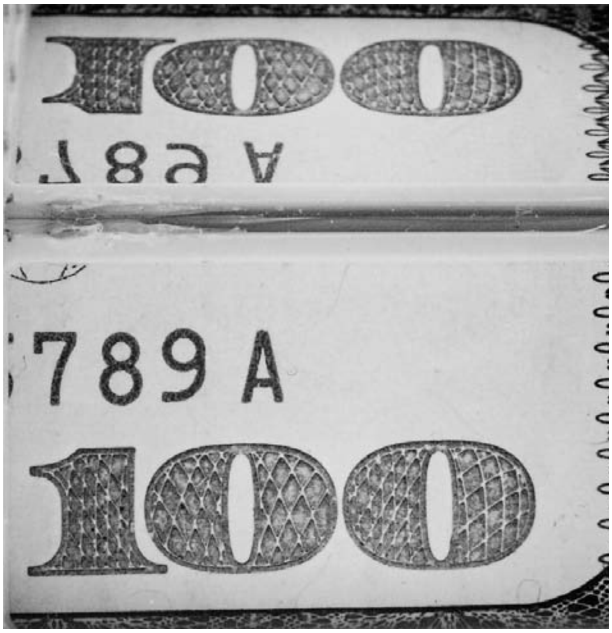


Figure 5 (see color plate 27) Green-to-black color shifting ink on genuine US banknote. Photographed using an angled mirror: green image (below) at normal; black (above) at an oblique angle.

Applied features

In modern banknotes, the security of the design may be further enhanced by inclusion of reflective, holographic or other diffraction-based, image-shifting foils. Such foils may be used as security threads, added during the papermaking process, or added to the surface of the banknote. A number of variations of these foils are available, from three-dimensional holograms to pixel or line-based diffraction gratings. In all cases, the image rendered changes at different viewing angles. Since holographic images are widely available commercially, the clarity and complexity of the images are meant as critical elements to the security of this type of device.

Authentication

As with any production process, errors can and do occur. When these errors reach circulation, they can be misleading to both human and machine authentication. Partly for this reason, secure, modern banknotes are generally designed with a checklist of features and characteristics, beyond the intrinsic design elements, which can be verified to prove or disprove authenticity. The process of authentication then becomes a routine of examining each characteristic until assurance is reached that the sample is a genuine note. If it is not, then it is a counterfeit.

It is crucial to have a basic knowledge of the genuine, including natural variations, to authenticate a document. Natural variations occur in each step of the process, not only in the printing but also in the production of the substrate, resulting in deviations in feathering patterns and in watermark appearance. Familiarity with natural variations can be acquired only by examining a sizeable representative population of notes produced over a period of time. This knowledge provides the basis for differentiating the characteristics of counterfeits from natural variations in genuine notes.

Currency authentication in vending, exchange or banking machines makes use of different machine-readable characteristics of a note. Counterfeit notes produced specifically to defeat these machines may lack image quality. Conversely, counterfeits to be passed to an untrained person may be a good likeness of the genuine note, but lack the less obvious security or machine-readable features. Because of these ambiguities, counterfeit notes range from absurdly obvious to extremely deceptive; therefore, various levels of analyses are required, depending upon the level of simulation/duplication within the suspect note. Crude

counterfeit notes will require much less scrutiny than more sophisticated counterfeit notes or notes that are ultimately determined to be genuine.

Examination Techniques for Suspect Currency

Many of the examination techniques used on suspect currency are microscopic, typically requiring 7–40 × magnification for security features, printed images and paper surface evaluations, and 40–100 × power for paper fiber examinations. Various illumination sources are also necessary, including near infrared (700–1000 nm), short- and long-wave ultraviolet and the flexibility to use them in reflected and transmission modes. Various optical filters used in combination with these illumination sources and detectors can be used to differentiate inks. Specialized equipment like microspectrophotometers and micro-Fourier transform infrared (FTIR) spectrometer provide physical and chemical information without the alteration or destruction of the sample. Other methods, such as UV-vis and IR spectroscopy and thin-layer chromatography, are necessary to examine the chemical components of the note.

During the course of examination, the analyst may find that the equipment or expertise needed to perform certain examinations is not available. In this situation, approaching another forensic laboratory, equipped with the necessary resources, may yield vital investigative or forensic information.

Paper/substrate

The foundation of any banknote is the substance on which the image is printed. Whether it is paper, polymer, or a combination of both, its properties and characteristics can have a great influence in the detection of a counterfeit. Because of the obvious role the substrate plays in the banknote, its evaluation is a logical place to start the examination. Most currency paper has features that can be examined physically and are often unique to the banknotes of the issuing country.

Physical examination *Standard light* Many of the properties of the paper/substrate can be observed under standard reflective white light illumination. Such features as the surface finish (topography), color and presence of overt security features can easily be viewed and compared.

Transmitted light Many security features, such as watermarks and embedded threads, are designed to be observable in transmitted light and unobservable in reflected light (**Fig. 6–9**).

Alternate light sources

- *UV* Many security features incorporated into the currency paper are designed to react with alternate light sources such as long- and short-wave UV light. Certain security fibers, threads and planchettes fluoresce when excited by UV radiation (**Figs 8**, **10** and **11**). UV illumination is also useful for revealing evidence of simulated features, such as printed watermarks, which may absorb or reflect the light differently from the genuine paper substrate.

Most commercial paper manufacturers increase the whiteness of their paper by the addition of UV-fluorescent optical brighteners. Conversely, genuine banknote paper is made without UV-reactive brighteners and, therefore, exhibits a dull UV response. However, it should be noted that genuine currency paper can become fluorescent by post-issuance contact with substances, such as laundry detergent, which contain optical brighteners.

X-ray Transmitted X-rays are useful for capturing differences in paper thickness, allowing visualization of commercial watermarks such as company logos, text or other images, which, if found on a suspect note, would confirm it as counterfeit. Additionally, X-raying a suspect note can confirm that a security watermark is a printed simulation rather than an actual watermark. Further, the small density difference created by the paper machine's forming wire can be imaged by this technique, allowing for comparison to the genuine.

Microscopic examination Microscopic examination is a powerful tool in the physical analysis of the suspect banknote, allowing the composition and security features of the paper or substrate to be analyzed. The analysis of the fiber content of the paper is conducted by sampling the paper and repulping it. The fibers are then stained and examined microscopically. The fiber species are identified by their morphology, and the paper manufacturing process can be classified by its reaction to various stains.

Security features Under magnification, security features like fibers, threads and planchettes can be examined for their physical characteristics, such as color and morphology. Microscopic examination can reveal whether these security devices were embedded in the substrate or simulated by printing or gluing them to the paper (**Figs 12** and **13**). In cases where the

Figure 6 Genuine US $100 Federal Reserve note (FRN) watermark.

Figure 7 Counterfeit US $100 FRN watermark: simulated by offset printing.

genuine security features are embedded in the substrate, further comparison will often reveal differences between the morphology, color or composition between a genuine and counterfeit feature.

Chemical examination Most currency papers incorporate fillers and sizing not generally used in commercial paper; therefore, chemical analysis of these additives can often differentiate between the counterfeit and genuine papers.

Spectrometric analysis of suspect polymeric banknotes may provide the necessary information to determine the note to be counterfeit.

Bleached genuine Because of the unique characteristics of currency paper, many counterfeiters remove the ink ('bleach') in an attempt to create more deceptive counterfeit notes. The counterfeiter may use low-denomination genuine notes from the target country, bleach the paper, and print the image of a higher denomination. Since the paper is genuine, the paper-based security features will also be genuine and an examination of these features will not alert the examiner to the fact the banknote is counterfeit. Alternatively, the counterfeiter may choose to use a foreign banknote in order to reduce expenditure. An examination of such a note would reveal that the genuine security features of the paper are not consistent with those used in the target country's genuine currency.

Bleached notes may exhibit characteristics of the process. The bleaching process usually involves some chemical and direct physical eradication of the image, resulting in partial removal of the surface sizing and disturbance of the paper fibers. Although extensive wear of a genuine banknote also results in fiber disturbance, this type of wear is characterized by image degradation not present in the counterfeit. Depending upon the effectiveness of the bleaching process, residual images of the original banknote may be apparent.

Quite often the counterfeiter resizes the paper to obtain better printability. The new sizing is usually evident under magnification. The resulting paper often has different surface topography and color when compared to the original genuine paper.

Figure 8 Genuine US $100 FRN security thread.

Figure 9 Counterfeit US $100 FRN security thread: simulated by offset printing.

Printed image

In addition to substrate analysis, counterfeits can be identified through analysis of the printed image. The three-dimensional, fine-line detail created by the engraved intaglio plates of the genuine image cannot be easily duplicated except by the very skilled and determined counterfeiter. Because people are skilled at face recognition, the small deviations in the portraits on counterfeits compared to genuine notes are expected to be more noticeable than deviations in other portions of the image. The portrait therefore serves as a logical starting point for image evaluation.

Standard light The first step in the evaluation of image quality is the comparison of the suspect note to a genuine note under good illumination. Differences in the images must be evaluated to determine if they are associated with wear, the natural variation of the genuine, a production anomaly, or because the note is counterfeit.

Some design elements incorporated into the genuine image create distortions (moiré) in counterfeit notes produced on low-resolution scanners and copiers or with half-tone processes. These distortions are obvious under normal viewing conditions.

Other security features, like color-shifting inks or optically variable devices, should be evaluated using techniques appropriate for that feature. This may require altering the angle of observation and noting the change in image or color.

Color differentiation of the inks can be aided by spectral analysis.

Transmitted light Examination with transmitted light is used to identify misalignment of front-to-back registered features, which may be an indication of counterfeit.

Alternate light Components with specific photo-active responses within banknote inks are common. Absence of a response or an incorrect response may be an indication of a counterfeit note.

Microscopic examination *Security features* The low-resolving power of most copiers and printers

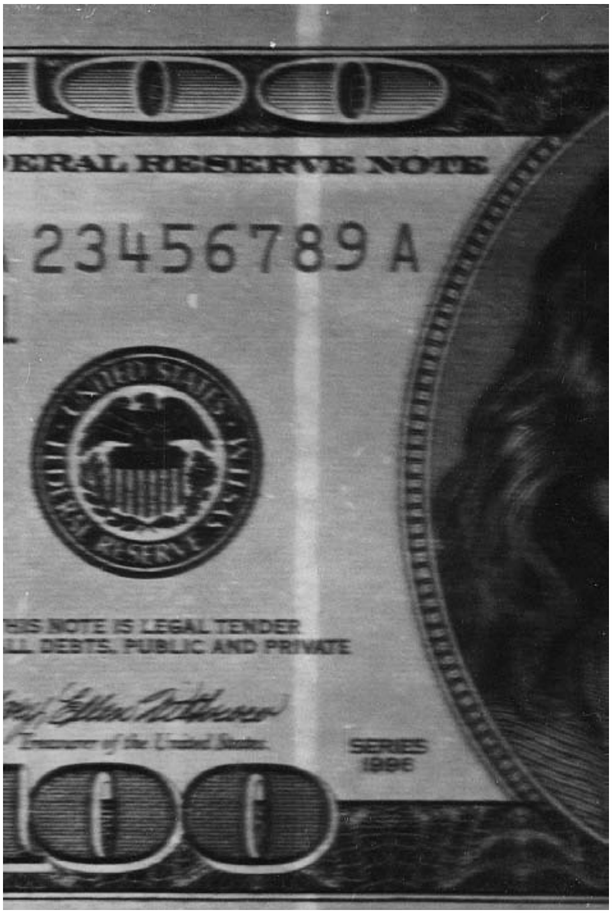


Figure 10 (see color plate 28) Fluorescence of genuine US $100 FRN security thread.

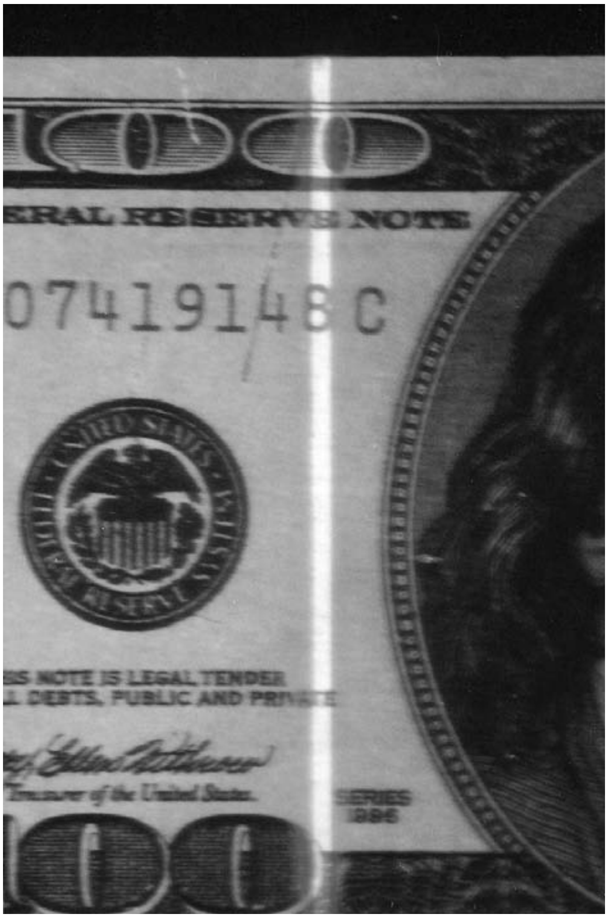


Figure 11 (see color plate 29) Fluorescence of counterfeit $100 FRN simulated security thread.

prevents faithful reproduction of small-scale details in the genuine notes, including microprinted text. If the microprinting is not legible, the note may be counterfeit (**Figs 14** and **15**).

Image deviation from genuine Some counterfeits are produced using methods and materials very similar to genuine (**Figs 16** and **17**). For these types of counterfeits, it may be easier to distinguish the counterfeit from the genuine based on differences in the images rather than on analyses from the materials used. Such cases require meticulous microscopic comparisons.

Printing process determination Images created by

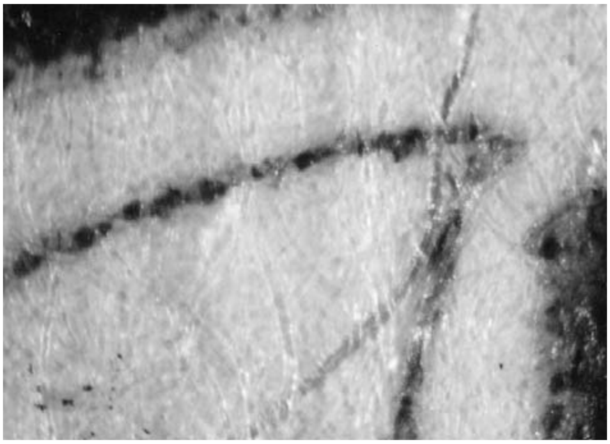


Figure 12 Genuine embedded security fibers.

Figure 13 Simulated security fibers: printed offset.

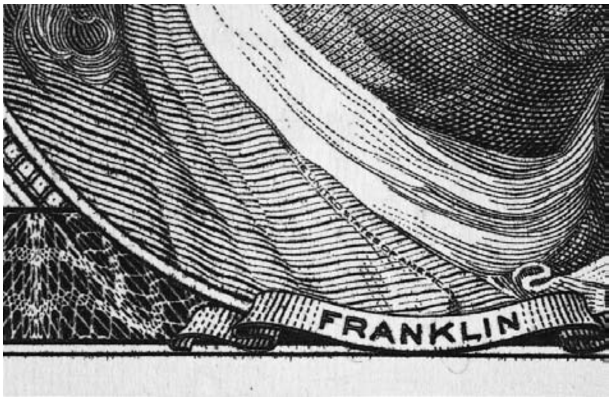


Figure 14 Microprinting on genuine US $100 FRN.

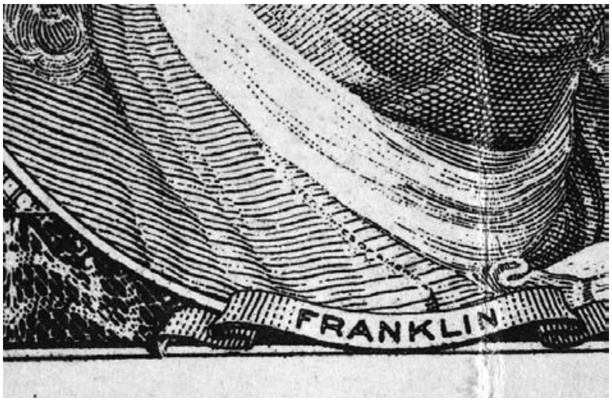


Figure 15 Microprinting on counterfeit US $100 FRN.

the various printing processes have corresponding properties characteristic of the physical or chemical means by which they were produced (**Table 1**). The most reliable tool for the determination of the printing process is microscopic analysis. In addition to intaglio, typographic, offset and screen printing (**Figs 18–21**), the following methods are also used to print counterfeits.

Halftone offset lithography The lithographic process can print only in a single density; it cannot generate gradations of densities to create tonal variations. To print tonal variations, the halftone process is used. This process translates images, with the use of a screen or computer, into a series of dots ranging in size and/or placement, giving the illusion of continuous tone.

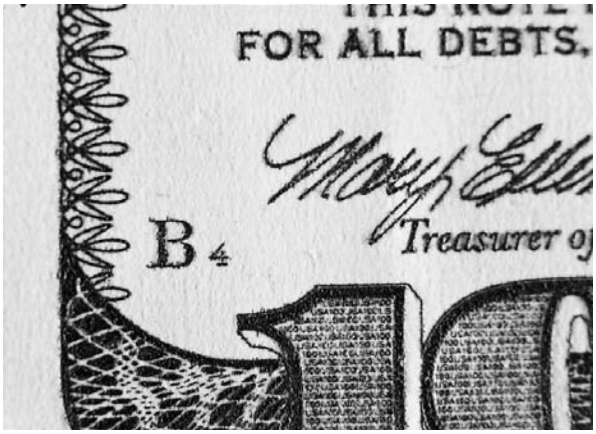


Figure 16 Genuine US $100 FRN.

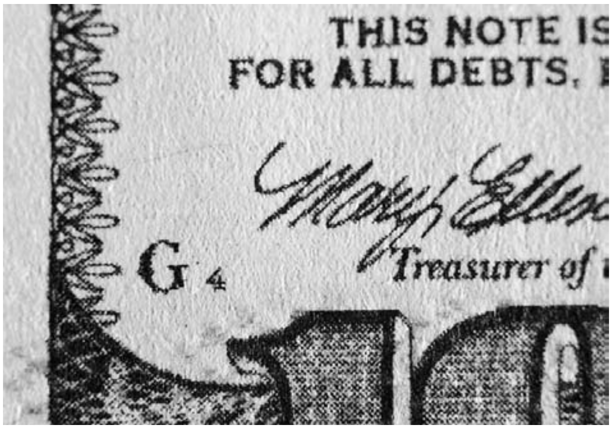


Figure 17 Counterfeit US $100 FRN: printed halftone offset.

Figure 18 Counterfeit US $100 FRN: printed intaglio.

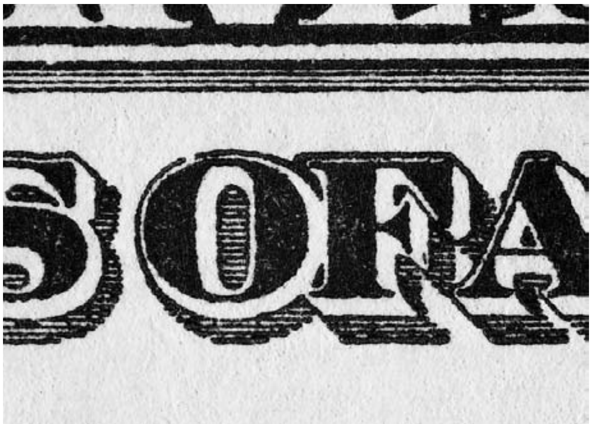


Figure 19 Counterfeit US $100 FRN: printed typographic.

Table 1 Printing process determination by printing process characteristics

Printing process	*Image*	*Characteristics*
Typographic (letterpress)	Imbedded printed image	• Ink is thicker near edges of image, creating an outline of the imaged area • Paper is embossed with the image
Intaglio	Raised printed image	• Feathering at edges of image • Paper in the image area is in relief
Screen	Raised printed image	• Fabric pattern
Electrophotographic	Raised printed image	• Dry toner – small plastic particles melted on to the paper surface
Offset lithography	Planographic/flat printed image	• Even ink coverage, sharp edges
Dry offset	Flat printed image	• Ink is generally thicker near edges of image, creating an outline of the imaged area
Inkjet	Planographic/flat printed image	• Undefined image edges • Ink is absorbed into the paper fibers

Full/process color halftone Process color separation and printing is a method used in the commercial printing industry to translate the various shades and colors captured in color photography into a halftone print consisting of only four colors: cyan, yellow, magenta and black. Although the resulting print consists of only four colors, it creates the illusion of reproducing the wide range of colours contained in the original (**Figs 22** and **23**).

Electrophotographic Electrophotographic processes that use dry toner (e.g. most copying machines

Figure 20 Counterfeit US $100 FRN: printed offset.

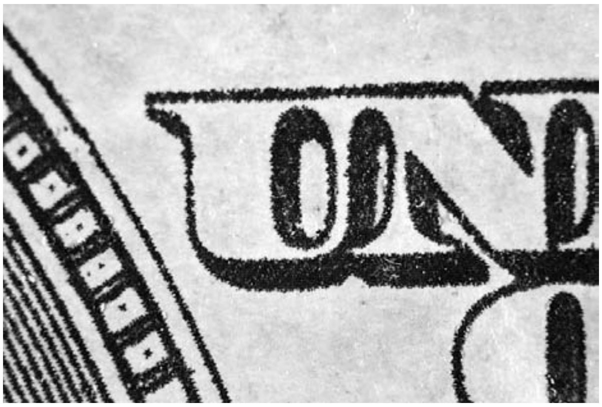


Figure 21 Counterfeit US $100 FRN: printed screen.

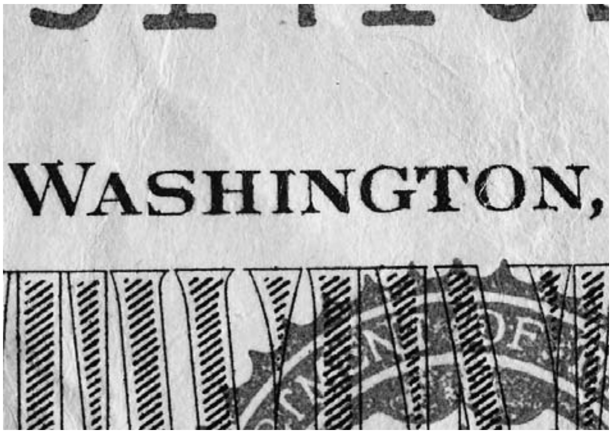


Figure 22 (see color plate 30) Genuine US $20 FRN.

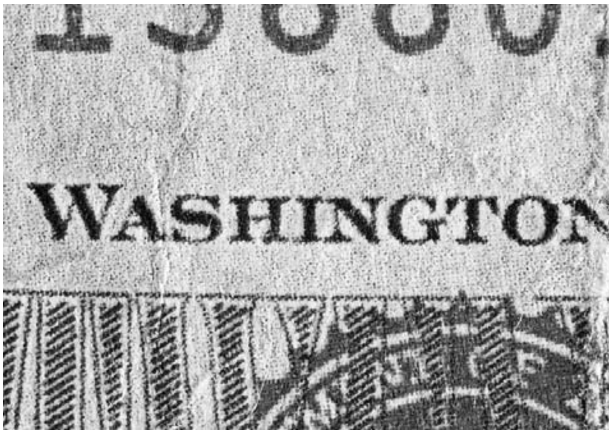


Figure 23 (see color plate 31) Counterfeit US $20 FRN: full color halftone offset.

and laser printers) result in images composed of tiny dots of fused plastic particles. Microscopically, the toner looks like plastic melted on to the surface of the paper (**Fig. 24**).

Inkjet Microscopically, these images are composed of ink spots that have been absorbed into the paper and appear flat. If two or more colours are printed close together, mixing of the inks might be apparent (**Fig. 25**).

Chemical examination Spectrometric, elemental and chromatographic analyses of the pigments, dyes and vehicles provide further avenues for comparison.

Investigative Information from Counterfeit

All aspects of the components and processes can provide valuable investigative information. This information may lead to the discovery of the counterfeiting operation.

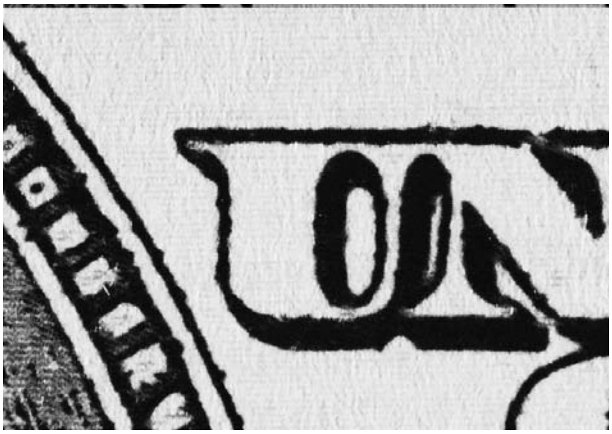


Figure 24 (see color plate 32) Counterfeit US $100 FRN: toner full color copier/printer.

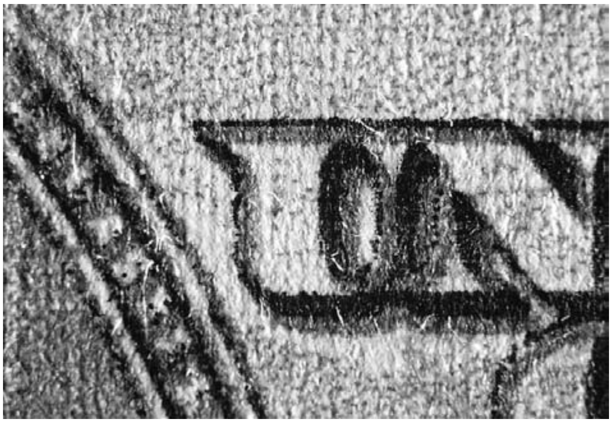


Figure 25 (see color plate 33) Counterfeit US $100 FRN: inkjet full color copier/printer.

Paper

Information such as the identity of the commercial paper based on identification of a watermark or identity of the country of origin by the paper's security features may provide a starting point for an investigation. The paper analysis may also provide valuable evidential corroboration with other counterfeits or with seized paper from a suspected counterfeiting operation.

Image comparison

The examination and analysis of the image not only provides printing process and equipment information, but can also supply the means to associate the counterfeit note with the components of its production. These include the original genuine note used as a pattern for the counterfeiting as well as any photographic negatives, printing plates or digital images used in the process. These components can be associated with the counterfeit based on the uniqueness of each genuine note and the transference of these unique characteristics to the images used to create the counterfeit. Image comparisons may also enable portions of different counterfeit notes, such as the front- or back-plate images, to be linked.

Inks Investigative and corroborative information may be obtained through the use of comparison libraries to identify the manufacturer of an ink used to produce a counterfeit note.

See also: **Analytical Techniques**: Microscopy. **Forgery and Fraud**: Payment Cards, **Document Analysis**: Analytical Methods; Forgery/Counterfeits; Ink Analysis.

Further Reading

Brunelle RL and Reed WR (1991) *Forensic Examination of Ink and Paper*. Springfield, IL: Thomas.

Carpenter CH, Leney L, Core HA, Cote WA Jr. and Day AC (1963) *Papermaking Fibers*. Technical publication 74: New York: State University College of Forestry at Syracuse University.

Church SE (October 1994) Security printing and specialty papers. *International Printing and Graphic Arts Conference*, Technical Section. 18–20 October. Montreal: Canada.

Committee on Next Generation Currency Design, Nationals Materials Advisory Board, Commission on Engineering and Technical Systems, National Research Council (1993) *Counterfeit Deterrent Features for the Next-Generation Currency Design*. Publication NMAB-472. Washington, DC: National Academy Press.

Durst SJ (1978) *History of the Bureau of Engraving and Printing*. Washington, DC: Sandford J Durst.

Eldred NR (1992) *Chemistry for the Graphic Arts*. Pittsburgh: Graphic Arts Technical Foundation.

Hessler G (1983) *The Comprehensive Catalog of U.S. Paper Money*. Port Clinton, OH: BNR Press.
Schafrik RE and Church SE (1995) Protecting the greenback. *Scientific American* vol. 273(1): July 40–46.
Stevenson DL (1994) *Handbook of Printing Processes*. Pittsburgh: Graphic Arts Technical Foundation.
van Renesse RL (ed) (1988) *Optical Document Security*. Boston, MA: Artech House.
Weatherford J (1997) *The History of Money*. New York: Crown.

Payment Cards

R N Morris, Ronald N Morris & Associates, Inc., Springfield, VA, Canada

Copywright © 2000 Academic Press

doi:10.1006/rwfs.2000.0447

Fraud: the Extent of the Problem

Payment card fraud is big business and hundreds of millions of dollars are lost every year to this criminal activity. The exact dollar amount of the losses depends upon the type of fraudulent activity and the country or region of the world where the fraud occurs. Since card programs – MasterCard, Visa, etc. – use different accounting methods to calculate losses, it is difficult to determine exactly how much the industry loses to fraud.

In economically developed countries, payment card fraud is more of a problem than in the European eastern block and other economically developing countries. In the developed countries, the use of payment cards represents a significant portion of each country's payment system. With the increased usage of cards in these countries, there has also been an increase in payment card fraud.

Payment card fraud will increase in all countries as the card mix available to consumers expands, from credit cards to check cards, cash cards, debit cards, etc. In developing countries payment card fraud and the resulting losses associated with their use will be of increasing importance. If the dollar losses are significant enough, they could affect the economic health and prosperity of the people in the affected countries. The goal of the payment card industry is to see that this does not happen and to reduce the level of fraud in the economically developed countries.

Payment card programs spend millions of dollars each year on card security features, fraud prevention training, investigators, etc. As new card security features are developed and implemented, criminals immediately try to compromise them. A couple of examples will help illustrate this: the use of the hologram or optical variable device (OVD) and the encoded magnetic stripe. Shortly after cards with OVDs were issued to the public, criminals compromised them by using substitute OVD image foils on their counterfeit cards. In most instances the image foils they used had images that did not resemble a genuine image. Many of them were image foils on children's stickers.

The criminal has only one question to answer when making a counterfeit card: will this card be a passable product, such that somebody will accept it and pay out cash or sell expensive goods because they think it is genuine? If the card is accepted at the point of sale, the criminal has succeeded.

In the 1980s organized criminal gangs in Southeast Asia had people who would supply them with holographically produced image foils with images closely resembling genuine image foils. The quality of the foils and the counterfeit cards they were on posed a significant threat to the payment card industry. Visual inspection of the card by the merchant was not sufficient.

The card payment system was in a transition period at this time, in part because of the quality of the Asian card. The magnetic stripe was becoming more important as a means of verifying the card and for payment authorization and reconciliation. Because the industry was becoming more dependent on the encoded magnetic stripe, new security methods and procedures were developed to insure its integrity.

The magnetic stripe is encoded with account information embossed on the card, and security information not embossed on the card. The magnetic stripe is discussed in more detail later in this article. It was not long after the magnetic stripe was introduced that criminals obtained magnetic stripe readers and encoders, and determined how the stripe, the information on it and the associated payment system worked. Electronic card fraud had begun and now magnetic stripe encoding, reencoding and skimming is a significant problem to the industry.

Criminals compromise card security features as soon as the features appear on issued cards. The card programs continue to research and develop new card security features to keep the products' vulnerability to criminal activity as low as possible. It often seems like an endless succession of invent, implement, compromised, invent, implement, compromised ... The industry goal is to attempt to stay ahead of the criminals and to control losses resulting from their activities.

Types of Fraud

There are many types of payment card fraud. A few examples are:

- application fraud;
- unauthorized use of a card;
- use of another person's identity to obtain credit or to make counterfeit identification documents;
- fraudulent bankruptcy schemes;
- counterfeit cards, white plastic and altered genuine cards;
- merchant bust-out and collusion fraud;
- magnetic stripe skimming and reencoding of card stripes.
- telephone and cell-phone card fraud;
- fraud schemes that use Internet web sites offering products or services, but requiring the buyer to give a credit history or credit card number first.

A detailed discussion of each of these is not the purpose of this article. However, where payment cards are widely accepted, in the United States and Canada, for example, the use of counterfeit and altered cards seems easier for the criminal. This is in spite of card program and issuer rules, neural network software systems and regulations governing merchant responsibility. The main issue for the industry is how to control losses.

Payment card fraud is cyclical. Counterfeit card use may be high for a few years and then drop off; application fraud may take over, only to be replaced by an increase in merchant fraud; skimming is currently the explosive area. Why this occurs is unclear.

Organized Efforts at Fraud

Payment card fraud is a worldwide problem perpetrated by criminals acting alone or in organized groups. The organized criminal gangs are usually international in scope. They have sources to supply them with materials to make counterfeit cards, security features like OVDs and special embosser punch and die sets, and a distribution network for completed or partially completed cards. There is also a network of people to use the cards wherever shipped, and to funnel money back to the criminal leadership for use in other criminal enterprises.

In the early 1990s Detective Chief Inspector John Newton of the Metropolitan Police, London, carried out a worldwide study of organized Southeast Asian groups active in payment card fraud at that time. One of his conclusions was: 'Despite its undoubted success the card payment system does have weaknesses, which are exploited by organized criminals. Counterfeiters, in particular, rely upon the almost universal acceptability of payment cards to make their criminal enterprises extremely profitable. However, the industry and law enforcement agencies are not standing still and allowing criminals to plunder the system without fear of disruption and arrest.' Newton found that counterfeit payment cards made in Southeast Asia were being distributed to, and used in, the United States, Canada, Europe and many other countries around the world. He also learned that many of the people arrested in one region of the world were rearrested in other regions, or they were in some way connected. The cards they were arrested with were often linked together forensically, showing that the card components had common source characteristics.

Some of the activities engaged in by these organized groups are:

- counterfeiting of cards and card security features;
- false identification documents;
- the use of payment card fraud to fund other criminal activities.

The information to make cards and the equipment and items needed for counterfeiting them are available from many different sources. The Internet is one such source of information. Some Internet sites contain programs to generate possible active account numbers and information on how to encode a stripe and what information is encoded on it. Magnetic stripe readers and encoders, embossers, etc. can be purchased from catalogs, over the Internet, value-added distributors, and even as second-hand office equipment. Frequently, equipment purchased in one part of the world is small enough to be hand-carried or shipped to another part of the world, where it is used to personalize cards.

It is not necessary to have the card to commit fraud: having only the account information is sufficient. Some gangs have members who just obtain account information that is used on their counterfeit and altered cards. The account information necessary to carry out a transaction is called an 'access device'. When a criminal has the access device, he or she has access to the full line of credit or services available up to the limit on that account. Access device fraud is prosecuted in the United States under local and state criminal statutes, and Federal Statute 1029, which makes it a Federal crime as well.

Genuine Cards

Payment cards are made in secure, authorized manufacturing facilities. Each manufacturer has some proprietary procedures or equipment they use in making their cards, but the result is nearly the same. How is a genuine card made? The following general discussion covers most manufacturers:

- The card issuer receives permission from the card program to issue cards. Only banks are authorized to issue payment cards. Cards can be cobranded, where a nonissuer like Shell Oil Company can issue a card in conjunction with Chase Manhattan Bank.
- The issuer creates artwork for the card or creates the artwork in conjunction with the card manufacturer's graphic arts section. The card program has very specific specifications concerning card real estate and how much of it can be used by the issuer and how much by the program. The issuer and manufacturer must conform to these standards.
- As the artwork is prepared, the printing process(es) for each component part of the card is determined. It is doubtful that the whole card will be printed using only one printing process. Genuine cards are usually printed using a combination of screen and offset printing.
- The printing plates for each phase of the printing process are made from the artwork.
- The inks are selected, mixed and color tested.
- The card is not printed as a single card but as one of a number of cards on a large sheet of plastic. In some printing plants, printing as many as 80 card images on one sheet of plastic is a common practice.
- After printing, the sheet of card images is laminated to protect the printed material from damage during card usage.
- The magnetic stripe is applied to the card as part of the lamination process.
- The individual cards are die cut from the sheet, and the signature panel and OVD are attached to the individual card.
- The card is then personalized, embossed, encoded and sent to the cardholder.

The Card in Greater Detail

A payment card is a sandwich of plastic. It has a white core, covered by a translucent overlay on the front and back (**Fig. 1A**). The white core plastic contains the printed material, program and issuer names, logos, ownership information on the back of the card, etc., which is covered by the overlay for protection. The OVD is hot stamped to the overlay on the front of the card (Fig. 1B). Not all card programs use an OVD, so the presence of this feature is dependent upon the card program and issuer. The back of the card contains a magnetic stripe and signature panel (Fig. 1C). The encoded information on the magnetic stripe is for payment authorization and reconciliation process, and the signature panel is for the cardholder to sign his or her name after receiving the card.

A card is a combination of numerous parts and technologies. The following discussion explains how the card should be viewed and something about each part.

The plastic

The plastic is polyvinyl chloride (PVC). The nominal size of the card is 3.370 inches (85.60 mm) wide, 2.125 inches (53.98 mm) in height, 0.030 ± 0.003 inches (0.76 ± 0.076 mm) thick, and the corners are rounded with a radius of 0.125 ± 0.012 inches (3.175 ± 0.305 mm). The card is die cut from a larger sheet. Depending upon the card manufacturer, the core may be a solid sheet of white PVC or two thin sheets sandwiched together which are sealed during lamination. Typically, the core is 0.027 inches (0.686 mm) thick and the thickness of each overlay sheet is 0.0015 inches (0.0381 mm). This size card is described as an ID-1 card by International Standards Organization (ISO) Standard 7810.

Not all PVC is the same. For example, two PVC plastic manufacturers may use the same compounds in their plastic, but the printability characteristics and whiteness can be very different. Since the core is a vital part of the package, because all of the printing is done on its surfaces, printability is very important.

The composition of the translucent PVC overlay

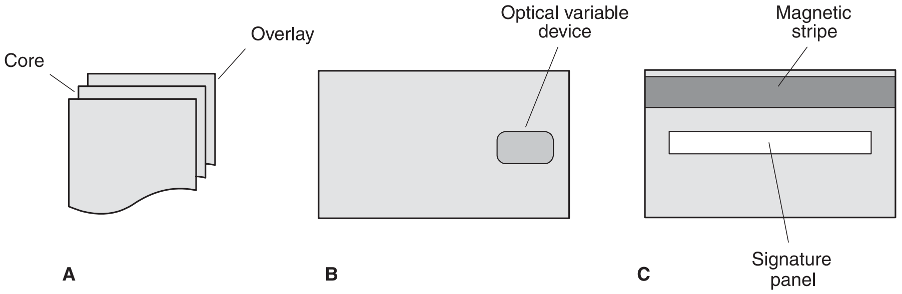


Figure 1 (A) A payment card is a sandwich of polyvinyl chloride. The core is white and is covered by translucent overlays. **(B)** Some card programs use an optical variable device (OVD) on their cards as a security feature. **(C)** The back of the card has a magnetic stripe and signature panel. See text for details.

material can also be different. The UV characteristics of the overlay are very critical because of UV fluorescent ink images printed on the core. The overlay must allow the UV light to pass through it and react with the UV printed ink image so it will fluoresce.

Ink

Both offset and screen inks are used to print on cards. UV curing inks or drying agents are added to the ink to expedite the drying process. In either case, it is critical that ink surfaces are dry before they make contact with another sheet of plastic.

A color shift occurs in the ink during lamination. To compensate for this color shift, a test is conducted to insure that the ink color after lamination is correct. The final ink color required is the one after lamination.

Signature panel

The signature panels are as varied as the card programs and issuers. Some are plain, offering little or no security, others are sensitive to solvents, or have complex printed backgrounds or other high security features. The signature panel may be one of the following:

- A preprinted hot stamp panel attached to the card.
- Several layers of screen-printed ink applied to the card, consisting of a base layer with printed information over it identifying the card program.
- Special paper or printed material and inks sensitive to solvents that will react with the stripe if an attempt is made to alter the written signature.
- A high security printed panel with a variety of printing traps to prevent duplication by using a scanner.

The panel may contain the account number indent printed within it.

Magnetic stripe

A magnetic stripe payment card can have three encoded tracks of information, depending upon the width of the magnetic medium. If the stripe width is approximately 7/16 inches (11.11 mm), three tracks of information can be encoded, and when the stripe width is approximately 5/16 inches (7.94 mm), two tracks of information can be encoded. Other technologies, such as watermark magnetic, etc., are beyond the scope of this article.

Encoding consists of a series of zeros and ones, a binary format, where each number, letter or character is a unique combination of '0' and '1'. ISO Standard 7811-2 and 7811-4 establishes the number of bits per character. The standard also requires an odd parity for each character bit set, regardless of the character, so included in the character bit set is a parity bit to insure an odd parity for the set. Odd parity is achieved by adding a one or zero to a bit set, insuring it has more ones than zeros.

The following is a brief description of the characteristics of a magnetic stripe data format based on ISO standards.

Track 1 This is known as the International Air Transportation Association (IATI) track. The character configuration is seven bits per character, including the parity bit, and the maximum information content is 79 alphanumeric characters, which includes the start sentinel, end sentinel and longitudinal redundancy check character (LRC) at a recording density of 210 bits/25.4 mm per inch. The card magnetic stripe data format for track 1 is illustrated in **Fig. 2** and described below. The encoded information is subdivided into fields read from right to left, as if the reader were looking at the back of the card.

A. Starting clocking bits Before data is encoded a string of clocking bits is encoded and each bit equals zero. Clocking bits provide a timing mechanism for starting the reader synchronization process before the read head gets to the start sentinel.

B. Start sentinel The start sentinel (%) is a defined number of bits informing the magnetic stripe reader that the following encoded information will be data.

C. Format code The format code consists of two digits telling the magnetic stripe reader how to interpret the data encoded in the data field. ISO established a standard of F2F (frequency-double frequency) for payment card magnetic stripes.

D. Primary account number (PAN) The PAN field is the first data field read by the magnetic stripe reader and can have a maximum of 19 digits. The card program and issuer establish the PAN and its format in the data field. The format of the PAN is not the same in this field for all tracks on all payment cards.

E. Field separator Following the PAN is a field separator ({). Its purpose is to notify the magnetic stripe reader that all encoded data called for in the preceding field is encoded and a different information field is about to begin. On track 1, it signals the end of

K	J	I	H	G	F	E	D	C	B	A

Figure 2 The IATA track or track 1.

a numeric data field, the PAN, and the beginning of an alpha data field, the name of the card owner. On tracks 2 and 3, it signals the end of the PAN and the beginning of the next, numeric data field. The PAN field is frequently less than the allotted 19 digits. If the encoded number is 16 digits, the issuer or program may still require the use of the full 19-digit field by leaving the unused digits blank.

F. Cardholder name The next field on track 1 is the cardholder name, and its allotted size is a maximum of 26 alpha characters. How much of this field is used and the order in which the name is encoded (last name first, first name second, initials third, etc.) are established by the card program and/or issuer.

G. Field separator A field separator follows the card owner's name. Its function is the same as described in 'E' above.

H. Additional data and discretionary data field The next field is a numeric field containing encoded information, the expiration date, service code, pin offset, 'Card Verification Value', etc. It is in this field that the VISA, Card Verification Value (CVV) and MasterCard, Card Verification Character (CVC) are located. The CVV and CVC numbers are security features established by the programs to insure the integrity of the encoded data on the stripe. Each card program and/or issuer can dictate the information encoded in this field. The ISO Standard establishes the number of characters allotted to each specific entry. For example: the expiration date, four digits; the restriction or type, three digits; the pin offset or parameter – optional – five digits, etc. Some issuers require the expiration date and 'member since' or 'valid from' date. In this situation, the date field will be eight digits rather than four digits. To do this the program obtains a variance from the standard. The card program and issuer establish what numeric information is encoded in this field, what it means, and how it is used.

I. End sentinel The end sentinel (?) is encoded after the discretionary data field. The end sentinel informs the magnetic stripe reader that the data fields of the stripe are now encoded and no further data is encoded.

J. Longitudinal redundancy check character (LRC) The encoding of an F2F format for payment cards call for an odd parity on each track. Each character has its own parity bit already; the LCR insures that the string of bits also has an odd parity.

K. Ending clocking bits After the LRC there is another series of clocking bits to verify the timing of the synchronization process.

Track 2 This was developed by the American Bankers Association (ABA) for the automation of financial transactions and its encoded information is numeric only, limited to 40 characters, including the start sentinel, end sentinel and LRC. The character configuration is five bits per character, including the parity bit, at a recording density of 75 bits/25.4 mm per inch.

Track 2 is located below track one on the magnetic stripe (**Fig. 3**). There is a small separation between the two encoded tracks to prevent magnetic field interference during encoding and reading. Since a detailed description of each field for track 1 is given above, and the explanation is the same for track 2, only the function for each field is given:

A clocking bits;
B start sentinel – HEX B;
C primary account number (PAN);
D field separator – HEX D;
E additional data or discretionary data field;
F end sentinel – HEX F;
G longitudinal redundancy check character (LRC);
H ending clocking bits.

The hexadecimal (HEX) is a numeric system based on powers of 16. Valid hex digits range from 0 to 9 and A to F, where A is 10, B is 11, . . . F is 15. In the above, HEX-B would be the encoded binary equivalent of '11', HEX-D is the binary equivalent of '13', etc.

Track 3 This was developed by the thrift industry and is a numeric data-only track with 107 characters, which includes the start sentinel, end sentinel and LRC (**Fig. 4**). The encoded density of track 3 is 210/25.4 mm bits per inch with five bits per character. The information encoded in the use and security data and additional data fields is updated after each trans-

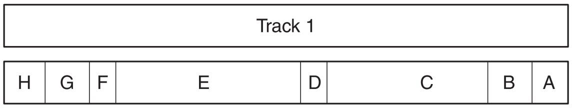


Figure 3 The ABA track or track 2.

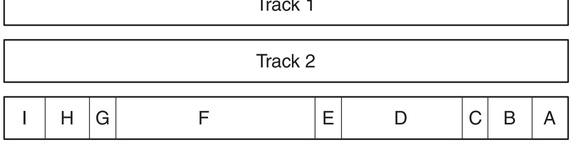


Figure 4 Track 3 was developed by the thrift industry and contains information that is read and updated after the transaction has occurred.

action, allowing a card encoded on track 3 to be used in online and offline systems. Since a detailed description of each field for track 1 is given above, only the function will be given for each field in track 3, as they, too, overlap:

A clocking bits;
B start sentinel – HEX B;
C format code;
D primary account number (PAN);
E field separator – HEX D;
F use and security data, and additional data field;
G end sentinel – HEX F;
H longitudinal redundancy check character (LRC);
I ending clocking bits.

Optical variable device (OVD)

Holograms have been used on payment cards for many years. The OVD foil is hot stamped on top of the front overlay of the card. Some European cards also have OVD foils hot stamped on the back of the card. Not all card programs require the use of holograms. The most popular ones requiring an OVD are MasterCard, Visa, JCB, Novus, and Europay, etc.

A hologram is traditionally defined as a three-dimensional image foil. However, not all holograms are three-dimensional. There are two-dimensional and three-dimensional foils. Some companies have trademark names for their OVD products, such as Gyrogram, Optoseal, Keingram, Exelgram, etc. Although there are these and many others, they all have one thing in common: a passable product of these OVDs can be used in their place on counterfeit documents.

Probably the best book on OVDs as a security device is *Optical Document Security* (see Further Reading). Having a copy of this resource in a technical library on questioned documents is essential. The reader is referred to this text for a more comprehensive study of OVDs and optical document security.

Embossing

Embossing is the production of the three-dimensional characters rising above the front surface, and recessed in the back of, the card. There are many different type/styles available for embossing cards. The ISO Standard 7811-1 mandates the type/style for the account number as a Farrington 7B OCR type. The remaining material, alphanumeric, on the card can be any type design as long as its size conforms to that allocated by the ISO standard. Some of the designs used over the years are Standard Goth, Simplex, Bradma, etc.

Embossers and embossing have unique characteristics that make them identifiable. The examination of embosser type is a three-dimensional study: in **Fig. 5**, the 'X' coordinate is the width of the typeface; the 'Y' coordinate its height, and 'Z' is the depth of the impression in the plastic.

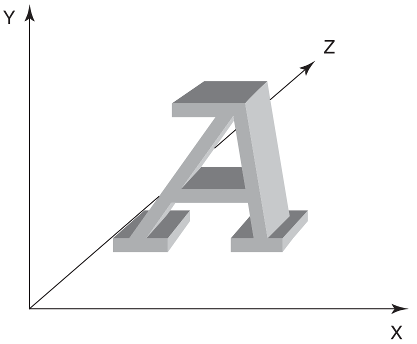


Figure 5 The examination of embosser type is a three-dimensional study. See text for details.

Figures 6 and **7** show an embossing punch and die describing each of its surfaces. Each surface shown can leave a record of its presence on the plastic overlay surface of the card. These defects, together with the class characteristics of the punch and dies, and the machine collectively, are what make embossing identifiable.

When the card is examined, it is important to observe the presence of any defects on the overlay of the card **Fig. 8**.

Printed material on the core plastic, such as the text or planchettes added to the background printing, make it difficult to focus on the top of the overlay. The interference caused by this background material is referred to as optical noise. Since the overlay is only about 0.0015 inches (0.031 mm) thick, it is difficult to focus on its top surface and ignore the optical noise under it on the core.

To illuminate the top surface properly, the angle of

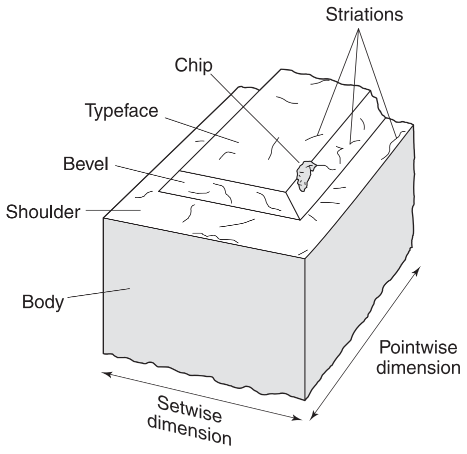


Figure 6 The parts of the embossing punch. This punch has a chip in the typeface and bevel, and striations on the typeface, bevel and shoulder. These individual characteristic defects are recorded on the surface of the plastic during the embossing process.

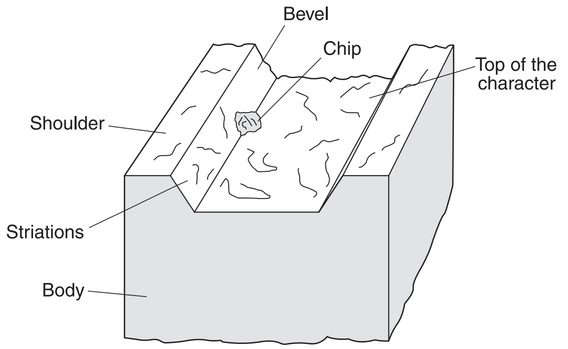


Figure 7 Defects (a chip and striations) in an embossing die.

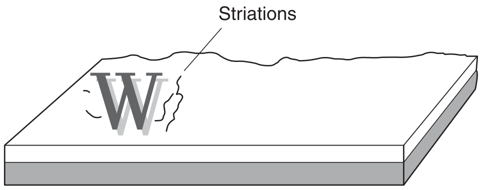


Figure 8 Surface defects on the overlay of an embossed 'W'. These defects may be found on the front and back of the card, either on or near the embossed character or inside it.

the light must be very small, as shown in **Fig. 9**. Use of a stereomicroscope and fiberoptic light source is best for this examination. Varying the angle of the card and light to the objective in small increments allows the examiner to see the greatest overlay surface detail possible. The purpose of this examination is to reflect light off the surface of the overlay, because this is the surface where the striations and defects are located. The use of a comparison microscope is essential when comparing the striations on separate cards or a card with the punch and die from a suspected embossing machine.

Two other methods used to assist in the examination of the card overlay surface are the vacuum deposition of a thin, opaque, layer of metal over the surface of the card where the embossing is located. The evaporation of this thin film is applied to the card surface before the embosser examination begins. This thin metallic film covers the surface of the overlay where the embossing is located and acts as a first surface mirror, highlighting the defects. Because this thin film is opaque, the problem of optical noise caused by the background on the core is overcome. A second method is the use of Mikrosil casting material. While this is a good material to show chips and larger defects in the plastic overlay surface, it does not seem to work as well for very fine microscopic detail defects.

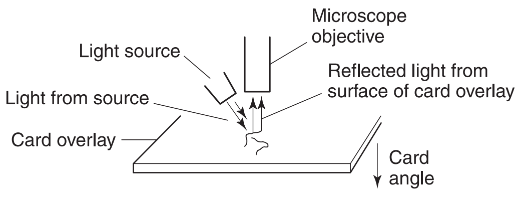


Figure 9 The relative angle of the light source to the card and the microscope objective for an examination of the overlay surface.

Counterfeit Cards

The counterfeiter has one objective: to produce a passable product. To be successful, the product only needs to pass the first line of inspection, the point of sale. Counterfeit currency must look and feel like genuine currency, while payment cards and identification documents must only look as if they are real.

When counterfeiting currency, it is important to have a piece of genuine currency from which to model the counterfeit. Since most people are familiar with currency, it is also important for the counterfeiter to copy its characteristics with attention to detail. Counterfeit currency is made using standard printing methods – offset, screen, intaglio and typographic – or a genuine piece of currency is copied on a color copier. Currency is not as complex a document for forensic examination as a payment card. The pictorial appearance and tactile qualities of genuine currency are well established and rarely changed. People know what a $20.00 US Federal Reserve note looks and feels like.

Counterfeit payment cards are more complex and varied than counterfeit currency, so the card counterfeiter has more of a challenge. There are hundreds of card programs and tens of thousands of issuers, and each one issues a number of different cards. Except for the program logo, the balance of the card real estate is very different for each card, often even from the same issuer. It is no wonder that merchants and point of sale personnel are not sure whether a payment card is genuine.

The number of printing processes available to the card counterfeiter is more numerous than those used by currency counterfeiters. Some of the processes used to print counterfeit cards are: offset, rainbow, screen, hot stamping, typographic or letterpress, thermal mass transfer, thermal dye diffusion, computer printers using transparencies that are then bonded to a plastic card, laser engraving, etc. Frequently, multiple printing processes are used to make a single counterfeit card. Linkage of cards based on the printing process used to print each area of the card is critical. Counterfeiters can, and do, share negatives and computer image software. They can, and do, make changes in the images they acquire, or they use a different combination of images and different printing process to print those images. That is why it is critical

to determine the printing processes used to print each card. Effective card linkage is based on common information on the counterfeit cards, the printing processes used, the common characteristics of the OVD and signature panel, and the fabrication techniques used to make the card and its component parts.

Further, the card OVD can be any of the following types: two-dimensional and three dimensional holograms, hot stamped images on diffraction foils, computer generated images using commercially available silver and gold paper, etc. Image analysis and linkage of OVDs is critical to establishing a common source of these devices. When examining hot stamped images on diffraction foil, it should be remembered that two similar images might have been made using different hot stamp dies. The details of the images must be examined carefully and be in complete agreement before concluding that two or more of them have a common source. The same basic principle also applies to the other card components. If all of the components of two cards agree, then it can be safely said that the cards come from a common source. If one component differs, then it can only be said that all of the components of the card agree except that one which has a different source from its counterpart on the other card(s).

The type of signature panel, or the printing methods used to make it, are very important in linking cards. The method of printing the panel or applying it to the card must be determined.

The fabrication techniques used to make the counterfeit card must be determined, in addition to the encoding of the magnetic stripe. It is not just what information is encoded on the stripe but an analysis of that encoding to determine something about the characteristics of the encoder that is used that is important. It is the combination of the different technologies used to make a counterfeit card and their forensic examination that make it a more complex document to examine than a piece of counterfeit currency.

Because there are so many different issuers of payment cards, it is impossible for a merchant or bank teller to know what each issued card is supposed to look like. A single issuer can issue several different cards authorized by more than one card program. For example, the issuing bank may issue gold and platinum cards, individual and corporate or company cards, cobranded and executive cards, etc. as separate product lines. The only common characteristic these cards share are those features mandated by the card program(s) authorizing the issuance of a card with their logo, signature panel, OVD, etc.

Counterfeiters can take many more liberties with the balance of the material on the card. They can use the same issuer name and logo, front plate details and back plate details for both a MasterCard and VISA card. Genuine MasterCards and VISA cards will not have the same information on the front and back of the card. In fact, the issuer's name that counterfeiters choose does not even have to exist. Counterfeiters can design a MasterCard using their logo and the name of a fictitious bank. They can create a logo for the bank and use an account number from another bank that does exist. As long as the merchant takes the card, because it is a passable product, the fraud is committed. Further more, counterfeiters are not obliged to adhere to the same quality and physical standards of a genuine card manufacturer. The card manufacturer must make a product that will last for years; the counterfeiter's card need only last for a couple of passes.

The linking of counterfeit cards is critical to an investigation. The Forensic Document Examiner plays a major role in this area because of unique skills developed over years of examining documents. A card is another form of a document and is forensically linked with other cards, as outlined below.

Fabrication techniques

Counterfeiters fabricate counterfeit cards using the same general techniques and similar equipment as a genuine manufacturer. The differences between the two are in the way they make the cards. For example, the Southeast Asian cards referred to above were fabricated as individual cards, not a part of a large sheet of cards. The printing was on the front and back of the white core and the card. The individual card was covered front and back with a precut card size overlay, and the package of overlays and core were placed in a mold, where heat and pressure were applied to seal the overlay to the core. While the card was still warm, small pickers were used to remove the card from the mold. The fabrication techniques were the same for different card program cards. Although the printed material on the cards had nothing else in common, these cards were linked by the common characteristics resulting from their fabrication.

Embossing

Cards fabricated in different counterfeiting plants may be linked by common embosser characteristics. The techniques for examining and identifying embossed characteristics have been discussed above. The author has linked together cards, by their embossing, which were obtained in Hong Kong, the United States and Europe. Linking cards by the embosser or punch and die used to emboss material on the card is very effective.

One word of caution, not all of the embossed

material on a card has to have been embossed on the same embosser or with punch and dies on the same machine. An example will illustrate this point. In the Hong Kong case, cards were shipped with embossed security logos and no other embossing on the card. The other embossed information was being added to the cards at their final destination. The embossed logos were linked to each other even if the remaining embossed information was done on different machines and could not be linked.

Printing on the card

The card must be broken up into its component printed parts to properly link them together by printing. For example, a MasterCard logo has at least four parts: the line that defines the area around the logo and OVD; the left part of the globe; the right part of the globe; and the word 'MasterCard'. Each of these component parts of the logo can be made with different source images that are interchangeable. It has been the author's experience after examining thousands of counterfeit cards that these parts of the logo are frequently made by mixing and matching interchangeable images that may come from different sources. Therefore, any linking system must consider this possibility.

The same is true for the balance of the printed material on the front and back of the card. A subgroup of experts developed the Interpol Forensic Classification System for Payment Cards, which is the most comprehensive forensic system for linking cards based on their components. Printing is one of the major areas of the Interpol classification system. The group did not address linkage of embossing. It was felt that such linkage was best left to the local laboratory forensic document examiner, as embossers were portable, and based on collective experience, international linkage of card embossing was rare. The system is designed to link cards from the same and different programs, different kinds of cards, printed material on card stock, OVDs and signature panels.

Magnetic stripe

It is not as yet possible to say with certainty that two cards were encoded on the same encoder. In certain instances it is possible to determine that two encoded magnetic stripes were encoded on different encoders, based on the characteristics of the encoder as recorded on the card's magnetic stripe. A special encoder analyzer is needed to perform this examination.

If the encoded information on the stripe of two different cards is the same, a linkage by encoded *information* is possible. This does not occur very frequently.

Optical variable device

Cards can be linked by their OVD. An image analysis of the OVD for common characteristics can reveal image features that can be used to link the image foils. Not all image foils have two- or three-dimensional images. Many counterfeit cards have reflective foils that are hot stamped to the card using a hot stamp die, and then using another hot stamp die with an image resembling a card program's genuine image. Sometimes the hot stamp image die and the foil are used as a single hot stamping, not a multiple hot stamping operation. Again, this demonstrates the importance of the fabrication technique of a card component. Linking OVDs is a significant part of the Interpol Forensic Classification System.

Where are Payment Cards Headed?

Payment cards are becoming even more complex. Chip cards are becoming more popular as the payment card system moves in that direction. Individual card security will depend more on cryptography, biometrics, digital systems and neural networks to insure the cardholder and the card in his or her possession go together, and to protect the card payment system.

Summary

Payment card fraud is an international problem. Organized gangs manufacture, distribute and use counterfeit cards around the world. Their cards are made using many of the same technologies as genuine cards. The forensic examination of counterfeit payment cards is performed by forensic document examiners because they have the unique skills to perform the required examinations involving printing processes, embosser identification and card linkage.

Cards are complex items and consist of components that have different common sources. Linking different components forensically assists investigators in suppressing the source of each component. One of the more important functions of the forensic document examiner is the linkage of counterfeit cards based on their common characteristics. Finding that counterfeit cards are linked by a single component does not mean that every other component on the cards are also linked; it only means that the one component has a common source. Linking each individual card component and the fabrication technique is necessary to reach the conclusion that the cards and their components come from common sources. The Interpol Payment Card Classification System is based on this concept, and it is designed to

link forensically individual card components, cards and fabrication techniques.

Payment card design by card programs is moving in new directions. Chip cards, cryptography, biometrics and neural networking are being used now, and will be used to a greater extent in the future to protect the payment card system. As card technology develops, the forensic examinations required will have to be further developed to keep up with an ever-changing payment card.

See also: **Document Analysis**: Analytical Methods Forgery/Counterfeits; Ink Analysis. **Forgery and Fraud**: Counterfeit Currency.

Further Reading

Morris RN (1998) *International Journal of Forensic Document Examiners* Apr./June, 4(2):128–133.

Newton J (1994) *Organized 'Plastic' Counterfeiting*. London: Home Office.

van Renesse RL (1988) *Optical Document Security*, 2nd edn. Boston, MA: Artech House.

Friction Ridge Impressions *see* **Cheiloscopy**. **Crime-scene Investigation and Examination**: Fingerprints. **Ear Prints**. **Fingerprints (Dactyloscopy)**: Automated Methods, including Criminal Record Administration; Chemistry of Print Residue; Identification and Classification; Sequential Treatment and Enhancement; Standards of Proof; Visualization. **Pattern Evidence**: Bare Footprint Marks.

Frye *see* **Legal Aspects of Forensic Science**.

G

Gender Determination *see* **Anthropology**: Morphological Age Estimation.

GEOGRAPHICAL PROFILING

M Godwin, Justice Center, University of Alaska, Anchorage, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0789

Overview

Over the last two decades numerous research studies using the theory of mental maps and home range have been applied to study criminals' spatial behavior. These studies reflect the importance of the journeys criminals habitually take around the areas close to their homes. The research also points out that such journeys provide criminals with information around which they are likely to plan their next crime; for instance, locations frequently passed by the criminal while traveling home, such as bars, shops and restaurants, are therefore proposed as defining the criminal's range and thus tuning his or her perceptions as to which areas are likely to be safe, both geographically and psychologically.

The concept of home range is based upon the notion that there is a geographical area around our homes in which we travel and use more regularly than areas further away. This area would typically include shopping areas, homes of friends and relatives and social activity areas. In 1913, it was suggested that home range can be thought of as a kind of cognitive or mental map; that is, images strongly related to residential location. Cognitive maps are representations of what is possible and where, built up through our daily activities and social transactions. As such, the mental maps we draw of an area change over time, and reflect how much time we spend in an area and the variability of our purposes for being there.

Each of these spatial approaches has a defining role in the specific aspect of geographical behavior investigated; for example, the location of the offender's home base; the location of the crime; and the relationship between the offender's home base and the crime scene. It is the latter area which has the most significance in the emerging field of geographical profiling. The identification of an offender's home base and the distances that the offender travels from home to commit crime is of value to police investigators who are investigating a series of crimes, for example, serial murder, rape or arson.

The Ecological Approach

Approaches to the study of geographical movements of criminals originated from what has become known as the ecological tradition, which developed in America between 1900 and through to the early 1970s. The ecological tradition is closely linked to a particular theoretical concept originally developed at the Chicago School of Sociology. Research in the Chicago School of Sociology in 1929, 1931 and 1969 exemplified the ecological principles. Using the ecological principles in 1942, Shaw and McKay confirmed the geographical coincidence of a number of social problems with high rates of delinquency and criminality. For example, they found that crime rates in American cities tended to be highest in the wealthier areas that were juxtaposed to poorer areas where there were identifiable concentrations of criminal residences. Shaw and McKay demonstrated the persistence of areas with a high rate of criminality over time and through changes in the ethnic make-up of these cities. Their research findings had a major impact on police procedures, and consequently have become accepted

as facts. Shaw and McKay's studies have influenced all subsequent sociological theory dealing with the geography of crime and delinquency.

However, the confusion in Shaw and McKay's ecological principles is that they assumed that the characteristics of an area with high proportion of criminal residents also identified and described the social characteristics of individuals who were likely to commit crimes. The attempt to use general data as an explanation for individual behavior has been termed the ecological fallacy by criminologists critical of the Chicago School of Sociology. As Brantingham and Brantingham pointed out in 1981, the problem comes when ecological data on the location of criminal residences is expected to answer questions about criminal motivations. The weakness of the ecological approach is that its findings are based on patterns of association between crime and its potential causes at an aggregate level and do not consider individualistic offending data. However, the ecological studies of the Chicago School of Sociology did draw attention to the potential of studying the spatial distribution of various urban, social and criminal indices.

Environmental Criminology

Deriving out of the ecological tradition, Brantingham and Brantingham in the late 1970s termed the phrase environmental criminology. Environmental criminology is concerned with criminal mobility and the relationship between offenders' home bases and their target areas. Environmental criminology attempts to predict the geographical area that an offender will victimize, based not on demographic features but on the individual's own mental image of the area. The Brantinghams proposed a theoretical spatial model for looking at journeys to crime as they occur in urban space. The Brantinghams' model uses concepts of opportunity and offender motivation together with the concepts of mobility and perception to predict the next likely target area and offender's resident. **Table 1** lists six major components that make-up the Brantinghams' hypothesized spatial model.

One element not shown in Table 1 is the spatial patterns of hunting and target selection. However, in an attempt to fill this void, the Brantinghams expanded their model using theoretical cases. The simplest case scenario and subsequently the foundation for current geographical profiling systems, is the basic search area for an individual offender. An example of the hypothesized search area can be seen in **Fig. 1**. The initial conditions in Fig. 1 are single offender, uniform distribution of potential targets and the offender based in a single home location.

The Brantinghams refer to the phenomenon of distance decay in describing the location of an offender's crimes. Briefly, distance decay refers to the reduction of activity or interaction as distance from the home increases. In the hypothesized case shown in Fig. 1, the expected range would be circular and most offenses would occur close to home, with the likelihood of an offense taking place in a particular location decreasing with distance from home.

The Brantinghams suggest offenders will have more cognizable mental maps about potential crime opportunities close to their home bases. They also proposed that offenders are more likely to be noticed and identified close to their home bases by other individuals who live in the same vicinity. Consequently, the Brantinghams argue that there would be a 'confront zone' directly around the offender's home base where little criminal activity would occur; this is referred to as a 'buffer zone.' For example, in 1980 it was suggested that it takes effort, time and money to overcome distance; these things are what have been termed the 'friction' of distance.

Table 1 Major elements of the Brantinghams' hypothesized spatial model

- Individuals exist who are motivated to commit specific offenses
- The environment emits many signals about its physical, spatial, cultural, legal and psychological characteristics
- The commission of an offense is the result of a multistaged decision process that seeks out and identifies a target or victim positioned in time and space
- As experiential knowledge grows, an individual learns which individual cues, clusters of cues and sequences of cues are associated with good victims or targets. These can be considered a template, which is used in victim selection
- An individual who is motivated to commit an offense uses cues from the environment to locate and identify targets
- This template becomes relatively fixed and influences future behavior

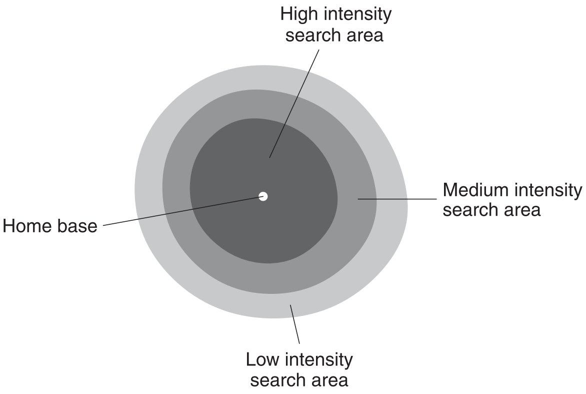


Figure 1 The Brantinghams' hypothetical offense area for a single offender. From Brantingham and Brantingham (1981).

In later research, the Brantinghams refined their hypothesized spatial model and proposed a complex search area for offenders. An example of this hypothesized model is shown in **Fig. 2**. The initial conditions allow for the fact that criminals, like noncriminals, are not tied just to geographical locations near their home base. Rather, offenders, like nonoffenders, go to work, shop and relax and the pathways between them all combine to form, what the Brantinghams termed, the individual's awareness space. Arguably, the Brantinghams' hypothesized spatial theories suggest that, given equally distributed opportunities, offenders will tend to offend within a minimum and maximum range of distance from their home, independent of direction and other physical or psychological constraints. Research that tends to echo this hypothesized spatial model was carried out in 1974; for example, it was found that in Cleveland, Ohio, property offenders tended to travel further from home than personal offenders did. Conversely, another study in 1982 found that black robbers and not burglars traveled longer distances to their crime sites. However, there are exceptions to these findings. For example, a 1969 study did not find any significant differences based on offense type and the distances between the offender's residence and crime location for 1960 Philadelphia data. Also, a study as far back as 1955 showed that 40% of all Houston homicides, including domestic homicide, between the years 1945 and 1949 occurred within one city block of the offender's residence.

Computerized Geographical Profiling

Following on from the work of Brantingham and Brantingham, Rossmo, calling his technique criminal geographical targeting (CGT), has combined concepts from environmental criminology with a mathematical model, based on a distance decay function, derived from the locations in which killers leave their victims' bodies, to indicate the area in which an offender may be living. The psychological principles on which Rossmo's work is based are not articulated in any detail but appear to derive from the postulate propounded by Brantingham and Brantingham's research suggesting that victims are probably spatially biased toward the offender's home base. This theory was illustrated in a study by the Brantinghams in 1981 in Washington, DC, where they found that offenders generally victimized areas they know best, concentrating on targets within their immediate environments and surrounding areas. This spatial bias is the proposed cause of a decay function, such that the further an offender is from home the less likely he or she is to commit an offense.

The reasons for the proposed decay proposed in Rossmo's criminal geographical targeting model are not exactly clear but appear to be based on the least-effort principle. As defined by one author in 1950, least-effort principle postulates that when multiple destinations of equal desirability are available, all else being equal, the closest one will be chosen. However, the notion that criminals choose targets in their confront zones is usually modified by two further considerations. One is most readily supported by the 1982 study referred to above, which found that specialist selectivity is reflected in the effort an offender puts into planning a crime. This choice would lead to more selective and more carefully planned crimes being committed further away from home. This theory has been supported by comparisons across different types of crime. For example, a series of studies in 1975 and 1976 found that the apparently more impulsive crimes of rape are committed nearer to home than robbery, and that armed robbers travel further on average than those who are not armed, and tend to net larger sums of money.

Rossmo incorporates another principle into his geographical profiling system, which has been put forward as a basis for crime locations: that there will be a tendency for serial killers to avoid committing crimes close to where they live, which was referred to earlier as a buffer zone. The proposed reason for this so-called 'confront zone' is so that criminals will avoid leaving incriminating evidence near to where they live. However, the evidence for this is sparse. For example, a 1997 study of 54 American serial killers found that offenders, on average, tended to make initial contact with their victims closer to home than the locations in which they eventually placed the bodies, which could suggest that a buffer zone is

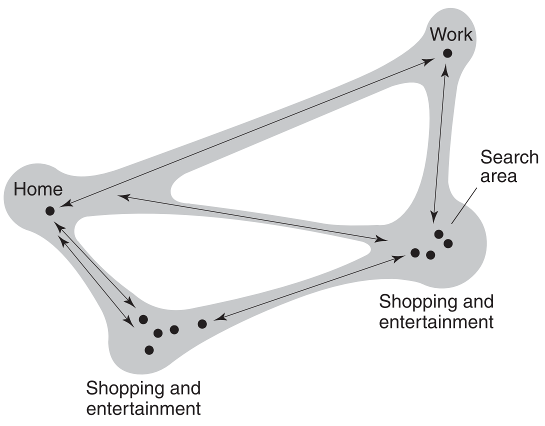


Figure 2 The Branginghams' hypothetical offense area for an individual offender, taking into account his or her action space. Arrows indicate movement paths. From Brantingham and Brantingham (1981).

highly probable for body disposal sites but not for the abduction sites. A study in 1995, for example, found no evidence for a buffer zone in a limited study of single rapists. Furthermore, the drawback to relying on the distance decay theory is that the actual distances proposed as buffer zones are often larger than would be consistent with leaving local clues.

A variety of geographical profiling processes have thus been outlined by those with an ecological and environmental criminology bias and these processes could be seen to be logically in conflict. One is a tendency to minimize effort close to home, which would predict that crimes are in a closely circumscribed area. A second is the tendency to keep a minimum distance away from home. These two processes combined would lead to the prediction of an optimum distance from home to all cases of a particular type of offense. However, the general finding is one of an aggregate decay of the frequency of crimes as their distances increase from home. These processes are derived from a consideration of instrumental crimes often with a tangible material benefit, such as theft or robbery. So, although they doubtless have relevance to geographical profiling, there are questions about how important emotional issues are ignored by such rational models. For example, two further complexities raise questions about the relevance of these rational processes to geographical profiling criminals. The present author has pointed out that there is typically more than one location involved in serial killers' activities. Besides the location of the site where the victim's body is dumped, there is also usually at least one other important site, the point at which the victim is first encountered. For example, in a series of murders, all three of the processes indicated above, least effort, buffer zone and decay function, would likely predict that the abduction and body dump sites would be close together. However, research on 54 serial killers by this author found that the locations where victims go missing were on average 1.5 miles (2.4 km) from the offenders' home bases compared with 5 miles (8 km) for the body dump locations. Furthermore, none of the three processes would lead to any predictions in the changes of the distance of the crimes from home over time.

Environmental Psychology

Following on from the environmental criminological theories proposed by Brantingham and Brantingham is the emerging field of environmental psychology. Environmental psychologists see the journey to crime as an expression of a complex interaction between the offender, his or her background characteristics, predispositions, knowledge and perceptions, and the location and type of target, in terms of perceived risks, rewards, opportunities and attractions. For example, in 1989 environmental psychologist, Canter, hypothesized that the actual nature of the location selected may be indicative of the purpose and experiences of the offender. Canter pointed out that there may be patterns of space use typical of different criminals, relating to where they are living at the time of their crimes.

Using an environmental psychology approach, Canter and his colleague in 1994 proposed research into the relationship that may exist between the criminal range and the location of the home base of serial sexual offenders. The proposed model of individual sexual offenders' spatial activity was based upon 45 British male sexual assaulters who had committed at least two assaults on strangers. The study hypothesized two general models to characterize the relationship between the home base and criminal area of offenders. The first assumption that was suggested regarded the geometry of a criminal's domain: that it would be an area defined by a circle around the offender's home, which Canter and his colleague defined as the commuter hypothesis. According to this theory, the area around the home and the area in which the crimes are committed are represented as circles. An example of the commuter hypothesis is shown in **Fig. 3**.

In describing the commuter process, Canter suggests that an offender travels from his or her home base into an area to offend. Central to this hypothesis is that there will be no clear relationship between the size or location of the criminal's domain and the distance it is from an offender's home. As such, the commuter hypothesis model proposes little or no overlap between these two areas, suggesting that the offender moves an appreciable distance to a district outside his home range to offend.

The second model proposed by Canter and his colleague is based upon the marauder hypothesis. An example of this hypothesized model is shown in **Fig. 4**. This theory argues that the offender's home base acts as a focal point for his or her crimes; it is suggested that the offender moves out and returns to the home base each time he or she commits a crime.

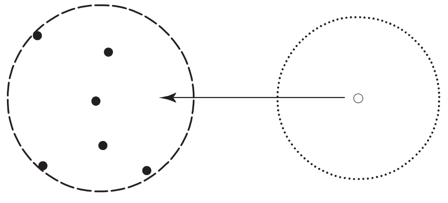


Figure 3 Commuter hypothesis. Open circle, home base; solid circle, offense. From Canter and Gregory (1994).

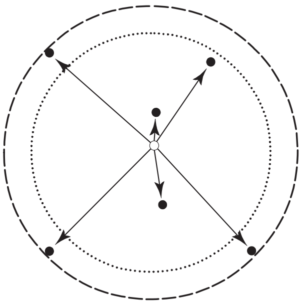


Figure 4 Marauder hypothesis. Open circle, home base; solid circle, offense. From Canter and Gregory (1994).

Arguably, this hypothesis would predict that, on average, the further the distance between crimes, the further the offender must be traveling from home to offend. However, research on 54 serial murderers found that the marauding process may be viable for victims' body dump locations but not for their abduction sites.

Canter's study found that 91% of the sample of offenders had all their crimes located within the circular region, which was termed the circle hypothesis. It was also found that 87% of the offenders had a base within the circle hypothesis prediction area. These results provided strong support for the general marauder hypothesis as being most applicable to this set of offenders; that is, that the majority of the sample had operated from a home base from which they traveled to commit their crimes.

Home Range and Directional Travel of Criminals

There is a further complexity about environmental criminology, Rossmo's geographical profiling model and the circle hypothesis processes to consider; that is, the hypothesized spatial models fail to take into account directional bias. For example, research in 1985 demonstrated that crime trips of suburban burglars are more likely to be skewed in the direction of nodal centers, such as workplaces and city central districts, and that crime trips were more likely to be wedge-shaped. Considering the plethora of research literature on the sectoral mental maps (directionality) of intraurban travel patterns and migration of people, the consideration of directionality in geographical profiling has been mute. One author argued that intraurban migration takes place in accordance with sectoral and directional biases in the urban spatial cognition. The findings suggested that intraurban migration patterns are sectorally biased towards the center business districts, in that they are wedge-shaped.

The theoretical analysis of people's bonds with the tangible surroundings of the home environs is found in several disciplines; for example, migration and shopping behavior studies. It has been pointed out that, through daily travel, the home environment becomes a unique place of familiar, known and predictable activities, people and physical elements, a focal point of one's experiential space. Thus, through habitual, focused and satisfying involvement in a residential locale, the tangible home area becomes an enduring symbol of self, of the continuity of one's experiences, and of that which is significant and valued by the inhabitant. The landscape around the home base may thus be hypothesized to provide criminals with those enduring symbolic experiences. If their crimes, as hypothesized, do indeed develop as an elaboration of their daily activities, rather than as some distinct work-life activity, then it would be predicted that the home would be geographically as well as symbolically central to their criminal activities.

Geographical Profiling the Angles Between Crimes Using Predator©

Predator is a geographical profiling system developed from research carried out by this author. The Predator system employs two location-allocation theories. Firstly, it employs a method that looks at the location of central facilities and the allocation of flows to them. Such a method projects, theoretically, the probable trips of the offender to the center. Secondly, Predator employs an analysis of a dispersing offender and reverses the directional assumption of the location-allocation model, while also keeping the assumptions of a monotonous plain and of optimization. Rather than assuming that the victim travels to the offender's home base, it assumes that the offender travels outward from his or her home, so the 'allocation of flows' is 'from' rather than 'to' the center. This process allows the criminal's spatial movements to be modeled from all directions; no matter if they are moving in or out.

Mapping the crime data involves recording each killer's crime locations and home base and their respective latitude and longitude geographical coordinates. Once this task is completed, all the geographical coordinates of latitude and longitude are converted into The Universal Transverse Mercator (UTM) grid system. Converting the latitude and longitude coordinates into the UTM coordinate system allows the data to be entered into the Predator geographical profiling system. Each distance is then entered separately into the Predator system. The

program then creates on the computer screen a scale of the actual crime area in miles or kilometers. For each event, the northing and easting UTM coordinate serves to express the unique, absolute location of each crime event. Equally important, the UTM coordinate system is impartial, and the values of locations are independently verifiable. The geographical program then plots each of the crime locations on the computer screen. Briefly, each case is analysed separately.

By way of example and demonstration of how offense angles can be calculated, **Fig. 5** shows the cognitive maps of four American serial killers that were calculated by the Predator system. Briefly, the data were collected from the Homicide Investigating Tracking System (HITS) based in Seattle, Washington. In instances where there was more than one home base, the mean of the angles between the crimes is reported. The finding that serial killers' spatial behavior has a tendency to be wedge-shaped has several implications for geographical profiling and predicting the likely home base area of a serial killer. For example, in the study of British sex offenders discussed earlier, the home base was found 71% of the time between the two furthest crime sites, as Canter's circle hypothesis suggests. However, there is still a rather substantially large search area for the police to investigate. If police were to rely on the wedge-shaped theory as a function in geographical profiling, the search area including the offender's home base would be considerably smaller. It is therefore argued that the wedge-shaped theory has heuristic value when developing geographical profiles of serial killers. For example, research on serial rape conducted in 1993 found that crime patterns had a distinct sense of directionality that could be described as a windshield wiper pattern.

In another study on directional bias, it has been suggested that the sectoral mental map may not be anchored on downtown, based on regular daily activities, but may instead focus upon recreation areas beyond the city which are visited weekly. To be sure, there is ample basis for the argument that criminals' spatial patterns depend on limited mental maps or mental images of the city and areas surrounding their homes. When they carry out their crimes, the locational choices are affected by reliable knowledge of the city or area, which forms wedge-shaped images that are shapes in focus for places close to home or other parts of the home sector, and blurry or blank for distant places, such as the other side of town. This paper certainly supports the view that investigative efforts would be greatly enhanced if geographical profiles took into consideration the directionality of crime locations.

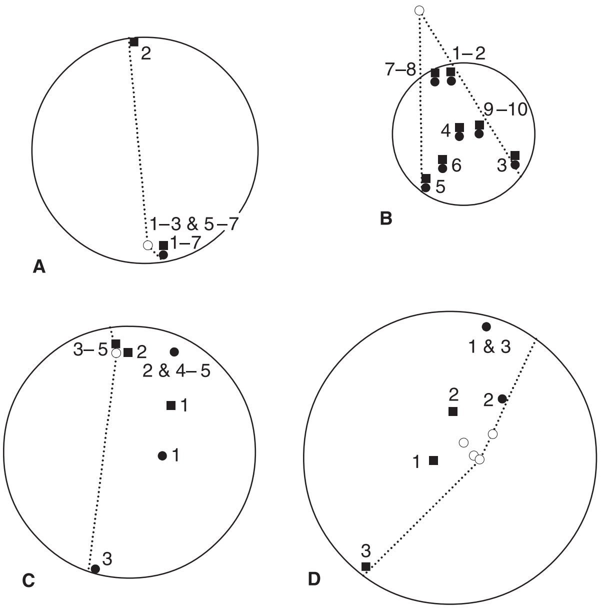


Figure 5 Cognitive maps of killers' movements between home bases, victims' abductions and body dump sites. **(A)** Offender 1: 1 inch = 1 mile (2.54 cm = 1.6 km); 130°. **(B)** Offender 2: 1 inch = 10 miles (2.54 cm = 16 km); 33°. **(C)** Offender 3: 1 inch = 20 miles (2.54 cm = 32 km); 246°. **(D)** Offender 4: 1 inch = 10 miles (2.54 cm = 16 km); 162°. ○, home base; solid circle, dump site; ■, abduction/last seen site; numbers represent order of crimes. Note: Difference in circle sizes is due to map magnification. © 1998 Predator.

Conclusions

This article has discussed theories and concepts associated with the geographical behavior of criminals, with particular focus on four principal approaches: ecological school, environmental criminology, environmental psychology and the wedge theory. As previously mentioned, research into geographical profiling originated from the ecological approach and was concerned with the identification of crime areas in relation to offender demographics rather than the location of specific criminal events.

The environmental criminology approach provided a new perspective on the relationship between an offender's home base area and the location of his or her offenses. Hence, it moved away from looking at the causes of crime and emphasized the importance of where victims are targeted and where crimes occur. Although the foundation on which environment criminology is built is entirely theoretical, it does suggest that criminals do have patterns in their spatial behavior, which could have implications for predicting future target areas. In a different vein, geographical profiling from the environmental psychological approach deals directly with the prediction of the likely behavior of the serial offender, as demonstrated in research on rape by Canter

and his colleague. Environmental psychology holds more promise than previous nonpsychological-based geographical profiling methods.

The article has presented evidence suggesting that criminals have limited spatial knowledge, and that these patterns appear to be wedge-shaped. Such findings are in accord with a perspective that sees journeys to crime growing out of the daily activities and contact patterns of the offender. For example, when serial killers make environmental choices involving locational considerations, their mental maps are used as preferences as to which areas to forage for potential victims and dispose of their bodies. The fact that serial killers' spatial behavior is sectorally biased suggests that police investigators are not dealing with economic killers, with perfect knowledge, but rather real individuals with imperfect knowledge and a predilection for satisfying behavior. Such a finding has direct implications for systems such as Rossmo's criminal geographical targeting system, leading to the hypothesis that his procedure might be more efficient if it considered the directionality of crimes.

Admittedly, the interpretations of this article are *ex post facto* and one has to speculate if there are additional features of the criminal's spatial behavior that can be predictive. Further research is needed to find out what additional psychological and environmental factors influence the geographical behavior of criminals.

See also: **Psychology and Psychiatry**: Psychology.

Further Reading

Adams J (1969) Directional bias in intra-urban migration. *Economic Geography* **45**:302–323.

Brantingham PJ (1978) A theoretical model of crime site selection. In: Krohn M and Akers RL(eds) *Crime, Law and Sanction*. Beverley Hills, CA: Sage.

Brantingham PJ and Brantingham PL (1981) *Environmental Criminology*. Beverley Hill, CA: Sage.

Bullock HA (1955) Urban homicide in theory and fact. *Journal of Criminal Law, Criminology and Police Science* **45**:565–575.

Canter D (1977) *The Psychology of Place*. London: Architectural Press.

Canter D (1989) Offender profiling. *The Psychologist*. **2**:12–16.

Canter D and Gregory A (1994) Identifying the residential location of rapists. *Journal of Forensic Sciences* **34**:169–175.

Canter D and Larkin P (1993) The environmental range of serial rapists. *Journal of Environmental Psychology* **13**:63–69.

Capone DL and Nicholas Jr W (1975) Crime and distance: an analysis of offender behaviour in space. *Proceedings of the Association of American Geographers*, pp. 45–49.

Capone DL and Nicholas Jr W (1976) Urban structure and criminal mobility. *American Behavior Scientist* **20**:199–213.

Davies A and Dale A (1995) *Locating the Stranger Rapist*. London: Police Research Group Special Interest Series Paper 3.

Down RM and Stea D (1973) *Image and Environment: Cognitive Mapping and Spatial Behavior*. Chicago: Aldine.

Feldman RM (1988) Psychological bonds with types of settlements: looking back to the future. *IAPS 10 Proceedings* **2**:335–342.

Godwin M (1998) Victim target networks as solvability factors in serial murder. *Social Behaviour and Personality: An International Journal* **26**:75–84.

Godwin M (1999) *Hunting Serial Predators: A Multivariate Classification Approach to Profiling Violent Behavior*. Boca Raton, FL: CRC Press.

Godwin M and Canter D (1997) Encounter and death: the spatial behaviour of US serial killers. *Policing: An International Journal of Police Management and Strategies* **20**:24–38.

Gold JR (1980) *An Introduction to Behavioral Geography*. London: Oxford University Press.

Johnston RJ (1971) Squatter settlements in South American cities. *Perspective* **8**:1–7.

Johnston RJ (1972) Activity spaces and residential preferences: some test of the hypothesis of sectoral mental maps. *Economic Geography* **48**:19–221.

Keppel R and Weis J (1994) Time and distance as solvability factors in murder cases. *Journal of Forensic Sciences* **39**:386–401.

Malmberg T (1980) *Human Territoriality: Survey of Behavioral Territories in Man with Preliminary Analysis of Discussion of Meaning*. Mouton: The Hague.

Mercer D (1971) Discretionary travel behavior and the urban mental map. *Australian Geographical Studies* **9**:133–143.

Pettiway LE (1982) Mobility of robbery and burglary offenders: ghetto and non-ghetto spaces. *Journal of Urban Affairs Quarterly* **18**:255–269.

Porteous J (1977) Mobility of robbery and burglary offenders: ghetto and nonghetto spaces. *Journal of Urban Affairs Quarterly* **18**:255–269.

Pyle GF (1974) *The Spatial Dynamics of Crime*. Department of Geography Research Paper No. 159. Chicago: University of Chicago.

Reboussin R, Warren J and Hazelwood R (1993) Mapless mapping in analyzing the spatial distribution of serial rapes. *Conference Proceedings: Workshop on Crime Analysis Through Computer Mapping*. Illinois Criminal Justice Information Authority, Chicago.

Rengert GF (1987) *The location of public facilities and crime*. Paper presented to the Academy of Criminal Justice Sciences, St Louis, MO.

Rengert GF and Wasilchick J (1985) *Suburban Burglary*. Springfield, IL: Thomas.

Rossmo DK (1993) Multivariate spatial profiles as a tool in crime investigation. *Conference Proceedings: Workshop on Crime Analysis through Computer Mapping*. Illinois Criminal Justice Information Authority: Chicago.

Rossmo DK (1995) Place, space, and police investigation: hunting serial violent criminals. *Crime Prevention Studies* **4**: Crime and Place.

Rossmo DK (1996) Geography, profiling, and predatory criminals. In: Holmes R and Holmes S (eds) *Profiling Violent Crimes: An Investigative Tool*, 2nd edn. Beverly Hills, CA: Sage.

Shaw CR and McKay HD (1942) *Juvenile Delinquency and Urban Areas*. Chicago: University of Chicago Press.

Stote R and Standing L (1995) Serial and multiple homicide: Is there an epidemic? *Social Behaviour and Personality* **23**:313–318.

Trowbridge C (1913) On fundamental methods of orientation and imaginary maps. *Science* **38**:990.

Turner S (1969) Delinquency and distance. In: Wolfgang ME and Sellin T (eds) *Delinquency: Selected Studies*, New York: Wiley.

Zipf G (1950) *The Principle of Least Effort*. Reading, MA: Addison Wesley.

Note: ©Predator geographical profiling system is copyrighted by Dr Maurice Godwin under the laws of the United States of America

Gun Powder *see* **Firearms**: Residues.

Guns *see* **Firearms**: Humane Killing Tools; Laboratory Analysis; Range; Residues; Weapons, Ammunitions and Penetration.

Gunshot Wounds *see* **Clinical Forensic Medicine**: Evaluation of Gunshot Wounds.

H

HAIR

Contents

Overview
Comparison: Microscopic
Comparison: Other
Comparison: Significance of Hair Evidence
Deoxyribonucleic Acid Typing
Hair Transfer, Persistence and Recovery
Identification of Human and Animal Hair

Overview

B D Gaudette, Royal Canadian Mounted Police, Ottawa, Canada

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0555

Introduction

Forensic hair examination is a unique field which was developed specifically for criminal investigations and which is not practiced in industrial or academic fields. It is, however, based on concepts and information developed in medicine, physical anthropology, wildlife studies, and the cosmetics industry. Knowledge of background information from these fields is essential to a forensic hair examiner. Of particular importance are the structure of hair, hair growth, the chemistry of hair, and other miscellaneous information on hair.

Hair Structure

Hair is a fibrous outgrowth from the skin of mammals which grows from papillae embedded in the bases of follicles situated in the dermis or true skin. Structurally, hair consists of three main parts: the cuticle, cortex, and medulla.

The cuticle

The cuticle, the outer scaly layer of the hair, is composed of flat overlapping cells or scales. Maintaining the hair's structural integrity, binding (preventing hair from fraying and breaking away in small fragments), and prevention of transfer of soluble substances from outside to inside the hair are the major functions of the cuticle.

The cortex

The cortex, which constitutes the major part of most hairs, consists of elongated cells fused together with intercellular binding material. Cortical cells consist of spindle-shaped fibrous structures called macrofilaments (macrofibrils), nuclear remnants, and pigment granules. Nuclear remnants are small, elongated cavities near the center of the cells.

Pigment granules are small, spherical particles 0.2–0.8 μm in diameter. They are composed of melanin, of which there are two types: eumelanin (brown–black) and pheomelanin (yellow–red).

Macrofibrils are mainly composed of microfibrils, which in turn are made up of protofibrils consisting of precise arrays of low sulphur proteins containing short sections of alpha helical proteins in a coiled formation.

Trapped within the cortex may be small air spaces called cortical fusi or vacuoles. As the hair shaft is pushed upward from the papilla, passage through the follicle transforms the cortical cells from an ovoid to an elongated spindle shape. The irregularly shaped tissue fluid-filled cavities that are carried with the cells later form the fusi. As the fusi are borne upward, they become longer and thinner and, as the cortex

dries out, they lose their tissue fluid and become air filled. Although in some hairs fusi can be found throughout the shaft, they are most common near the root. In general, the lighter the hair, the more fusi will be observable.

Because they are air filled, cortical fusi appear opaque under transmitted light. Though they are generally larger than pigment granules, they can sometimes be close in size, making distinction difficult. An examiner in doubt as to whether dark spots are pigment granules or cortical fusi can differentiate them by focusing up and down on the microscope. Fusi disappear from view, whereas pigment granules remain as out-of-focus dark spots. Alternatively, top lighting can be used. With this illumination, true fusi will appear bright.

The medulla

Most hairs usually have a central core, termed the medulla, which is composed of shrunken cells which may or may not contain pigment. The spaces between the medullary cells are usually filled with air, giving an opaque appearance under transmitted light. When cells are damaged or for some reason very shrunken, the mounting medium will fill the air spaces, leading to a translucent appearance. Depending on the type of hair and the individual, the size of the medulla is highly variable – constituting anything from 0 to 95% of the hair. In addition, the general appearance of the medulla varies enormously depending on the species from which the hair originated, the type of hair, body area of origin, and the portion of the hair shaft viewed.

Hair Growth

Growth of hair is achieved by the proliferation of cells in the matrix of the follicle and by their increase in volume as they move into the upper bulb. Since the cells are under pressure as they pass through the neck of the bulb, there must be a constraining mechanism to funnel them upward and to keep them from expanding laterally. Henle's layer of the outer root sheath is the first to become firmly keratinized and forms the first part of the funnel. The outer sheath, the vitreous membrane, and the connective tissue sheath give the funnel resiliency and firmness. Throughout its life, each hair follicle undergoes recurring cycles of active growth, regression, and rest. This is known as the hair cycle.

In most mammals, the hair cycle in each region of the body is synchronized, producing waves of growth activity which manifest themselves as seasonal variations. In humans (and guinea pigs), follicular activity is not synchronized, leading to what is termed a mosaic cycle, whereby neighboring follicles operate independently and are normally at different stages of the cycle.

There are three main stages in the hair cycle. Anagen is the phase of active growth; catagen the phase of regression; and telogen the resting stage. At the end of the active growth of the anagen phase, activity of the matrix ceases, and the hair moves into the catagen stage. There is an upward movement of the hair root which becomes bulb-shaped and surrounded by a capsule of partially keratinized cells, around which is an epithelial sac of epidermal cells. Once catagen is complete the hair simply rests in the telogen stage with no noticeable changes until anagen commences in the follicle. In the next phase, anagen, the follicle is reconstituted and a new hair grows up alongside the bulb hair. The bulb-shaped telogen hair is usually retained until anagen is well-advanced in the follicle, at which time it is shed. The human scalp contains compound follicles in which two, three, or even more hairs, each produced by a separate follicle, share a common orifice to the outside.

In alopecia (common baldness), normal follicular activity ceases. In telogen effluvium, psychological factors, disease, or pregnancy cause the cycle of many hairs to transform prematurely to the telogen stage and be shed. Side effects of cancer chemotherapy are the only known instance in which hair is shed from follicles which continue in the anagen phase.

In a healthy human scalp approximately 85–95% of the follicles are in anagen, 4–14% in telogen, and less than 1% in catagen. The average daily loss of human scalp hairs is about 75 to 100.

As seen in **Table 1**, the rate of growth of human hairs varies with location on the body. In addition, age, sex, race, hormones, and individual factors all influence the rate of hair growth. Although they appear to play little part in human hair growth, environmental influences have a considerable effect on animal hair growth.

The duration of the anagen phase ranges from 2 years and 2 months to 6 years and 6 months; catagen

Table 1 Rate of growth of hairs on various parts of the human body

Site	*Growth rate (mm day^{-1})*
Scalp	0.33–0.35
Beard	0.38
Axilla	0.30
Thigh	0.20
Eyebrow	0.16

lasts about 2 weeks, and telogen approximately 1 year. These figures can, however, be quite variable. After 3–7 years an individual will not have any of the same hairs on his or her scalp as he/she has today. This has important implications for hair comparisons in that if a period of more than about 5 years elapses between collection of the questioned and known samples, a meaningful comparison may not be possible. In other parts of the body, the hair cycle is considerably shorter than that for scalp hairs.

Chemistry of Hair

Human hair is composed of 65–95% protein, up to 32% water, 1–9% lipids, and less than 1% pigment and trace elements. The protein is composed of amino acids, 21 of which have been reported in human hair. The amino acid content of hair varies with the structural components of the hair and is influenced by genetics, weathering, cosmetic treatment, and diet. Water content of hair is dependent on relative humidity. The lipids in hair come principally from sebum and consist of free fatty acids and neutral fat (esters, glyceryl, wax, hydrocarbons, and alcohols). The chemical structure of hair creates unique physical characteristics. Strong disulfide bonds linking adjacent keratin chains produce a structure that is extremely resistant to chemical and biological degradation.

The chemical content of hair is influenced by genetic, environmental, and dietary factors. Chemical content can in turn influence structural and morphological characteristics. For example, it has been noted that the hair of persons suffering from extreme protein malnutrition is mechanically weaker, finer in diameter, sparser, and less pigmented than the hair of healthy individuals.

A large number of different trace elements have been reported in human hair. They can arise from metabolism, cosmetic preparations, or the external environment. Trace element content of hair is affected by personal factors (age, sex, diet, etc.), external factors (shampoos, air-borne contaminants, etc.), frequency of hair washing, the structure of the hair itself, and by sample treatment prior to analysis. Furthermore, hair trace element content varies with location on the body, from hair to hair on the same body area, from root to tip along the hair shaft, across the diameter of the hair, and according to the stage in the hair growth cycle. Because of such variation and sample treatment problems, trace element analysis of hair is not considered to be a reliable method of forensic hair comparison.

Miscellaneous Information on Hair

Animals have four types of hairs. These are the coarse outer guard hairs, the fine under fur hairs, intermediate hairs with characteristics in between the preceding two types, and vibrissae or tactile hairs such as whiskers. Similarly, there are four types of human hairs. The most obvious type is the terminal hairs such as scalp, pubic, beard, axillary, eyebrow, eyelash, and nasal hairs. Vellus hairs are fine, short, unmedullated hairs which cover the entire body. Again, there are intermediate hairs between these two types. Lanugo hairs are fine, soft, unpigmented, unmedullated hairs which cover the body in the prenatal period. Three classes of human terminal hairs have been identified: hair that is the same in both sexes (such as scalp hair); hair that is under hormonal influence and behaves as an ambosexual character (e.g. pubic and axillary hairs); and hair (such as beard hair) that serves as a true secondary sexual character.

The presumed functions of hair include decoration, protection against injury, increasing the perception of touch, insulation, acting as a filtering mechanism, and drawing away perspiration.

The adult human body has about 5 million hair follicles, the distribution of which over the body being somewhat sex and age related. Hair grows on all parts of the human body except the following: eyes, lips, areolae of nipples, umbilicus, anus, urogenital openings, soles, palms, finger tips and tips of toes.

The traditionally accepted average number of hairs on the scalp is 100 000–120 000, although a more recent estimate placed the number at 150 000–200 000. All estimates have been based on extrapolations of counts over small areas. No one has actually counted the number of hairs on a complete scalp. Further, there are no accurate data available on the average number of pubic hairs on the human body. In view of the implications to forensic science (such as collecting known hair samples), this lack of basic research is distressing.

Gross hair color is influenced by the microscopic color of individual hair shafts, the degree of medullation, cortical fusi, density of hair growth, hair diameter, cleanliness, and artificial treatment, as well as the angle and type of illumination. Microscopically, hair color is principally influenced by the type, density, and distribution of hair pigment.

Animals can show seasonal changes in hair color. The only comparable effect on human hair is the sun bleaching of the distal ends of scalp hairs often noticed in summer. Human hair color can change with age, generally becoming darker after childhood and then often gray to white in later life. The most noticeable change is graying, which is usually a mani-

festation of the aging process. The graying process results in a gradual dilution of pigment in affected hairs. In the early stages of this process, a full range of color from normal to white can be seen, both along individual hairs and from one hair to the next. Loss of color is associated with the decrease and eventual cessation of tyrosine activity in the lower bulb. Graying occurs at different rates in different areas of the body. Beard hair is usually the first to turn gray and body hair the last. The average age for the onset of graying in scalp hair is 34 in Caucasians, and by age 50, 50% of the population have at least 50% gray hairs. This grayness usually starts at the temples and gradually extends to the top of the head. Although graying is more obvious in dark-haired individuals, the rate at which hair pigmentary content is diminished is independent of the initial pigment concentration. Some diseases can temporarily affect hair color, and certain environmental influences can lead to detectable color changes in people with light colored hair. The hair of heavy smokers can be stained yellow. Exposure to high concentrations of copper can induce a greenish cast. Cobalt workers sometimes develop blue hair, and indigo handlers can have a deep blue tint in their hair.

The most important hair defects and diseases are monilithrix (beaded hair), pili torti (twisted hair), trichorexis nodosa (frayed nodes), trichorexis invaginata (bamboo hair), trichonodosis (knotted hair), and pili annulati (ringed hair). A detailed discussion of hair diseases and abnormalities is beyond the scope of this article. However, since they are occasionally encountered in forensic hair examinations, the reader is referred to the further reading list.

The disappearance of pigments from the hair shaft due to graying or disease creates tiny cavities in the shaft. The eventual compression of these cavities can cause a decrease in hair shaft diameter with age.

See also: **Hair**: Hair Transfer, Persistence and Recovery; Identification of Human and Animal Hair; Comparison: Microscopic; Comparison: Other; Comparison: Significance of Hair Evidence; Deoxyribonucleic Acid Typing.

Further Reading

Cottington E, Kissinger R and Tolgeyesi W (1977) Observations on female scalp hair population distribution and diameter. *Journal of the Society of Cosmetic Chemists* **28**:219–229.

Danforth CH (1939) Physiology of human hair. *Physiological Reviews* **19**:94–111

Ferriman D (1971) *Human Hair Growth in Health and Disease.* Springfield, IL: Charles C Thomas.

Harding H and Rogers G (1999) Physiology and growth of human hair. In: Robertson J (Ed.) *Forensic Examination of Hair.* London: Taylor and Francis.

Hausman LA (1932) The cortical fusi of mammalian hair shafts. *The American Naturalist* **66**:461–470.

Orfanos C and Happle R (eds) (1990) *Hair and Hair Diseases.* New York: Springer-Verlag.

Robbins C (1994) *Chemical and Physical Behaviour of Human Hair*, 3rd edn. New York: Springer-Verlag.

Robertson J (1987) Trace elements in human hair – significant forensic future? *Trends in Analytical Chemistry* **6**(**3**):65–69.

Rook A and Dawber R (1982) *Diseases of the Hair and Scalp.* Oxford: Blackwell Scientific Publications.

Saitoh M, Uzuka M and Sakamoto M (1970) Human hair cycle. *Journal of Investigative Dermatology* **54**:65–81.

Comparison: Microscopic

R E Bisbing, McCrone Associates Inc., Westmont, Illinois, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0560

Forensic Hair Comparison

Hairs, whether human or from a wild or domesticated animal, are often found associated with personal injury accidents and criminal activity. The determination of the source of a hair is often of interest to investigators and the principal means for determining its origin is by microscopical comparison. A comparison of all types of hairs is possible; although head and pubic hairs are most often encountered in criminal investigations, others can be compared using the same techniques.

History of forensic hair microscopy

Four centuries of microscopical study have described the morphological structure of hair, how populations of people differ, and how individuals differ (interindividual variation), and what variation can be expected within an individual (intraindividual variation). In the seventeenth and eighteenth centuries, hair was a ready-made specimen for Robert Hooke (1635–1703), Antony van Leeuwenhoek (1632–1723), and Henry Baker (1698–1774). In the nineteenth century, there were scattered reports of medical and forensic investigations into the use of hair as evidence. In the twentieth century, anthropologists studied hair as a means to learning more about human variation, and forensic scientists applied

microscopy to the identification and association of hair and used it as evidence. Hair is now well established as one of the many types of useful trace evidence and there is no reason to think that its use will not continue.

Morphological features for comparison

Forensic hair comparison is a comparative biological discipline grounded in microscopy, biology, anatomy, histology, and anthropology. In addition to biological processes, subsequent treatment of the hair and accidental changes due to the individual's environment create extra dimensions useful for comparison. Therefore, from the beginning, forensic hair examiners have concentrated on certain microscopic features, characteristics and traits for both human and animal hairs that have been shown, through collaborative research and collective experience, to be useful for comparison because these characteristics effectively distinguish one individual from others and one population from another. For human hairs, these characteristics can be broadly grouped into color, structure and treatment. Hair color, shaft, medulla and scales characterize animal hairs.

Color The color of the hair depends on its pigment, surface, transparency and reflectivity. A range of colors in the exemplar is common and a direct comparison is necessary between the questioned hair and the known hair in order to detect subtle differences. The amount and distribution of pigment within the cortex should be examined at high magnification (200–600 ×). Color is probably the most useful characteristic for comparison.

Structure The diameter (**Table 1**), medullation, cross-section, cortical fusi and spatial configuration reflect the hair's structure. The diameter can be measured and the medullation can be classified according to various schemes. The spatial configuration refers to the appearance of the hair as it lies, unconfined, on a surface and depends on the cross-sectional shape.

Treatment The tips, roots, color and length are all subject to treatment. Cut tips may be freshly cut, split, frayed or worn and the angle of cut may be significant. The shape of the root may not only indicate the method of hair removal, but abnormalities may provide valuable comparative data.

Bleached or dyed hair can usually be identified by a distinct demarcation near the proximal end between the treated and untreated portion of the shaft. The treated hair shaft sometimes shows signs of chemical wear; the cuticle is often damaged and cortical cells, separating under the chemical assault, may be distinct.

Microscopy

Light microscopy

Light microscopy is the use of any one or a combination of microscopical techniques including stereo, polarized, fluorescence and comparison microscopy.

As many comparisons may be solved quickly and inexpensively by light microscopy, it is the beginning point in nearly all forensic trace evidence analysis. In those cases where other techniques are indicated, light microscopy is used to prepare samples for further examination.

Illumination Good resolving power and optimum specimen contrast are prerequisites for forensic light microscopy. Though the optics (ocular, objectives and substage condenser) may be suitable, proper adjustment of the illumination is of paramount importance in order to see the details necessary for an adequate hair comparison.

Köhler illumination (August Köhler, 1866–1948) is the most useful illumination, and may be obtained with any lamp not fitted with ground glass. Diffuse illumination, so called because a ground glass is positioned between the lamp filament and the microscope condenser, seems easier to set up and gives illumination of high quality although of somewhat lower intensity. The illuminator and lamp filament need no adjustments, although a field diaphragm is helpful. Both types of illumination require that the

Table 1 Caucasian head hair diameter. Human hair shaft diameter measurements illustrate the range of values for many hair characteristics. These values were obtained from a random population of 50 individuals. The diameter distributions were often bimodal, probably reflecting the hair cross-sectional shape; and, although they overlapped, the distributions were unique for each individual. When many characteristics, each with their own distributions of values, are combined into a multidimensional space, the uniqueness of a representative sample of hairs from an individual can be imagined

Measurement	*Size* (μm)
Diameter range	13–132
Most common range	39–111
Average range	79.4
Grand mean	70
Smallest mean diameter	44.8
Largest mean diameter	90
Range of mean diameter	45.2
Smallest mode	39
Largest mode	104
Mode range	65

microscope substage condenser be focused on the specimen to give a full illuminating cone and allow the substage iris to be focused on the objective back focal plane. Defenders of the diffuser say they can obtain illumination in the field of view comparable to and, in fact, indistinguishable from Köhler. This may be true, but the only way to be sure the diffuse illumination is optimized for the hair comparison is to proceed through the steps necessary to achieve true Köhler illumination. If desired, the ground glass can be replaced in the light path.

In any case, Köhler illumination is not difficult for any microscope with a lamp that can be focused and centered. First, focus on a preparation using the high dry objective and be sure the rotating stage, if present, is centered with respect to the microscope's optical axis. Close the field diaphram and focus its image in the field of view by racking the substage condenser up or down. Center the focused, field diaphragm image by centering the substage condenser. Next, simply remove the ocular and look down the bodytube to see the objective back focal plane, and focus an image of the lamp filament in the objective back focal plane by moving the lamp bulb fixture along the lamp axis relative to the lamp condenser. Lateral movement of the lamp bulb itself centers the filament. Replace the ocular and, looking again at the preparation, adjust the substage iris to give optimum contrast and resolution. Each compound microscope needs to be adjusted similarly.

Calibration If linear measurements are to be made with the light microscope, the eyepiece micrometer must be calibrated for each objective. Likewise, the microscopes on each side of the comparison microscope must be matched to provide the same magnification and color balance for each objective.

The eyepiece micrometer is calibrated by comparing the ocular scale to be calibrated to a second scale on a stage micrometer slide having known dimension. The two scales are each carefully focused, the ocular scale by rotating the focusing top lens of the eyepiece and the stage scale by focusing the microscope itself. The scale images are then arranged parallel to each other and the size of each ocular division is determined by comparing with the length of the stage micrometer.

Color balance on both sides of the comparison microscope can be obtained with proper illumination, adjustment of the rheostat, new lamps, color-balancing filters or by illuminating both microscopes with a bifurcated fiber optic from a single lamp.

Cuticular scales The microscopical comparison of cuticular details illustrates the limits of a metrical approach to hair comparison and the effect of mounting media. The scale patterns of hairs are useful for distinguishing between broad classes of animals, and they provide some value for individualizing human hair.

Scales can be seen when mounted dry between glass slides; but, if required, better delineation of the cuticular surface is possible by preparing a clear cast, using fingernail polish for example. The scale casts are then viewed with a microscope with transmitted light. When hairs are mounted in a liquid with a refractive index near that of the hair (1.52–1.54), the scales disappear (unless stained) and cannot be used for identification or comparison.

The cuticular surface can be characterized by counting the number of scales per unit length or calculating a scale index defined as the ratio of the scale length to the hair shaft diameter. Scales can best be viewed with the aid of scanning electron microscopy because the small details are in three dimensions on the surface. Possibly, when more objective research data are compiled on the morphology of cuticular features using scanning electron microscopy and image analysis, their significance to species identification and hair comparisons will be enhanced.

Mountants Hairs can be viewed dry, simply mounted between glass slides; but, for a detailed comparison of the internal structures, hairs must be cleared with a liquid somewhat closer to the refractive index of the hair. The brand of mountant is not particularly important, but the refractive index and ease of use are (**Table 2**). Hairs mounted in water (n=1.33), for example, are too contrasty for an effective comparison of internal structures. There is too great a difference between the refractive index of the hair and the refractive index of the mounting medium. The mountant should have a refractive index near that of the hair cuticle and cortex, somewhere between 1.52 and 1.54. Mountants can either be volatile solvents, natural and synthetic oils, semisolid resins hardened by evaporation, thermoplastic solids or plastics cured by UV radiation.

Relatively permanent mounting media, such as Permount (Fisher Scientific Co.) or Cargille Meltmount 1.539, are frequently used in the United States to protect the integrity of the hair while at the same time clearing it sufficiently for microscopical comparison. If necessary, an easy way to remove the hairs from a semipermanent preparation is to first freeze the slide and then simply pick off the cover slip and the hair of interest. Sometimes the preparation in Permount needs to be soaked for some time in xylene or a Meltmount preparation can be warmed on a hot plate to remove the cover slip and hair.

Table 2 Mountants for microscopic hair comparison. The choice of mountant depends on what image is necessary for the analysis. If surface features are to be examined, a mountant with very low refractive index is used, like air. If internal details are to be examined, a mountant with a refractive index near that of the cuticle is chosen. Whether the hair must be used for another purpose or preserved indefinitely mounted on a slide determines if a liquid (solvent, oil or thermoplastic mountant) or plastic is used. The type and refractive index of several possible media are listed

Type	*Name/brand*	*Refractive index*
Dry	Air	1.00
Solvent	Water	1.33
	Toluene	1.49
	Xylene	1.50
	Orthodichlorobenzene	1.54
	Bromoform	1.60
Oils	Glycerine	1.47
	Mineral oil	1.47
	Immersion oil	1.515
	Cedar wood oil	1.52
	D C 710 silicone oil	1.53
	Clove oil	1.53
	Cargille Liquids	1.46–1.64
Semisolids	Protexx	1.49
	Permount	1.52
	Canada Balsam	1.54
	Histoclad	1.55
Thermoplastic	Cargille Meltmounts	1.539–1.704
	Aroclor 5442	1.660
UV plastics	Plastic Mount	1.49
	Norland optical #65	1.52
	Norland optical #68	1.54
	Norland optical #60	1.56

Stereomicroscopy The stereomicroscope, also known as a dissecting microscope, is the starting point for virtually every microscopical analysis. At low magnifications of approximately 5 × to 50 ×, the color, texture, structure and treatment of a sample may be observed and, in some cases, a complete identification can often be gained. Different colored backgrounds and different light sources should be used during initial comparisons of hairs. A black background often affects the appearance of the medulla and cortical fusi, for example. Stereomicroscopy is also used to prepare samples for the microscopical comparisons to follow.

Polarized light microscopy A polarized light microscope is based on a compound microscope and is the most useful of all light microscopy techniques, particularly for the identification and comparison of hairs, fibers and the small particles that might be adhering to them. Direct observation of hairs at magnifications of 40 × to 600 × provides morphological information and further manipulation of light provides optical information. Nearly every sample for analysis finds its way to the polarizing light microscope.

Because hair is long compared to its thickness, it is, in some ways, a difficult subject for thorough microscopical investigation. The study of important morphological details necessitates relatively high magnifications, and at these magnifications the depth of focus is less than the diameter of the hair, and the field of view limits observation to a short section of the shaft at any one time.

Fortunately, hair is anisotropic, and since birefringence increases with the degree of keratinization, the cortex is more birefringent than the medulla or cuticle. When a hair is examined between crossed polars, the background appears dark and the hair, if properly oriented, appears bright due to interference colors. The greater the birefringence or the longer the optical path length (thickness), the greater the ultimate phase differences and the more retardation, interference and orders of color. Therefore, portions of the hair, differing in either thickness or birefringence, appear in different colors, sometimes called optical staining.

In the examination of an unmedullated hair shaft, these colors indicate thickness. The optical pathlength through the narrower edges of a hair with an oval-shaped section is less than the pathlength through the thicker center. These color bands are especially useful to see structural variations, nodules, twists, constrictions or bandings.

The medulla, too, can be studied with the aid of the polarizing microscope. Under ordinary illumination, an air-filled medulla is a cylinder of low refractive index in a medium of higher refractive index and becomes a dispersing lens, so that the medulla appears deceptively black. Under the polarizing microscope, the medulla remains an isotropic inclusion in a birefringent matrix, and the retardation of polarized light passing through the portion of the hair containing the medulla is less than the retardation in surrounding parts. Hence the medulla appears in a different color from the surrounding cortex, and pigment granules (if present) remain dark, due to their specific absorption. In this way, the pigment content of the medulla is easily differentiated from empty space; or medullae, which are difficult to see in nearly opaque hairs, can sometimes be found.

Fluorescence microscopy Through the use of special attachments to a compound light microscope, the fluorescent nature of hairs may be determined, allowing hairs which are similar in normal or polarized light, but which have different fluorescence, to be distinguished. Fluorescence is most often used to

compare dyed fibers, but on occasion treatments or contamination of the hair may be detected with this technique. Because fluorescence is so valuable for fiber comparison, many comparison microscopes are equipped with the appropriate attachments so they are readily available for hair comparisons. Special stains may be used to treat materials on the surface to facilitate their detection by fluorescence microscopy. The color and intensity of fluorescence at different excitation wavelengths may provide important information for both the characterization and comparison of hairs.

Comparison microscopy Skillful microscopical technique when hairs are viewed simultaneously side-by-side on a comparison microscope furnishes the forensic hair examiner with a highly discriminating means to compare hair. The stereomicroscope allows gross observations of the hairs including trace evidence on the hair's surface. The wide range of traits in a large hair sample can be rapidly reviewed using the lower magnification and ample working space. The compound or polarizing microscope with its higher magnification and resolution delineates the hair's finer structural characteristics. The pigment, scale structure, cortical fusi and medulla can be scrutinized using the polarizing microscope and its accessories. Using a calibrated ocular micrometer, the necessary measurements can be made.

The final phase of a detailed microscopic comparison is a side-by-side examination of the known and questioned hair using a transmitted light comparison microscope where each of the microscopic features are juxtaposed and compared. An ideal comparison microscope would be of good quality with attachments for both polarized light and fluorescence microscopy.

Scanning electron microscopy

Scanning electron microscopy is valuable because it can image small structures with greater depth of field than can a light microscope, but it cannot see below the surface of the hair; therefore, for the purpose of hair comparison it is less reliable than the light microscope. The other reason scanning electron microscopy and elemental analysis are unreliable, either to prove identity or to eliminate a possible source, is that the techniques do not normally take into account variation along the length of the hair or variation from one hair to the next in the sample. The only reasonable way to compare hairs by electron microscopy is to make numerous measurements and observe numerous images at numerous points along numerous hairs from the root to the tip; therefore, it is generally unsuitable for making hair comparisons.

Infrared microspectroscopy

A Fourier transform infrared spectrometer attached to a light microscope allows analysis of samples down to about 10 μm in size. On occasion, a microscopical examination of either questioned or known hairs will reveal adherent material on the surface which will sometimes be useful for comparison. Infrared microspectroscopy can be used in addition to polarized light microscopy, fluorescence, and scanning electron microscopy for identification of these microtraces.

Visible light microspectrophotometry

Although color is very important to hair comparisons and with a stereo- or comparison microscope many different hues can be distinguished, microspectrophotometry of these differences has not been very successful. It seems that hair pigments, which are essentially all brown, produce flat and featureless absorption spectra in the visible region, 400–700 nm. A good way to define hair color has not yet been developed and it remains quite subjective.

Comparison of Hairs

Microscopical technique

The microscopy of hair is relatively simple. Using your hands and eyes and good microscopes, the hair is approached like a crime scene being investigated with the spiral technique. You survey the scene and begin a slow careful approach circling around in an ever-decreasing spiral until you have viewed all that there is to see. Look at the gross anatomy: how the hair sits on the surface, its shape, color, curl, length, etc. Next, circle around and get a little closer and look at the surface: microtraces might be there, the scales will be of interest including their size, shapes, and weathering along the length of the hair. Finally, peer inside, using a mountant near the refractive index of the cuticle and keratinous fibrils of the shaft, looking for: pigment particles, cortical fusi, ovoid bodies, the medulla, inner cuticular margin, etc. If you are looking at animal hairs look for: shaft shapes and sizes, uncut tips, color banding, cuticular patterns, unusual medullae, and root structures. If you must compare hairs to determine if they could have originated from the same individual, look at all these features again and determine whether they are similarly evident, in every detail, in both hairs.

Exemplars

Before a comparison can be attempted, the portions of the body from which the hair originated must be

determined. No comparison can be made without obtaining an exemplar from the homologous somatic region. Scalp hairs must be compared with scalp hairs, beard hair with beard hair, mane hair with mane hair, and dorsal guard hair with dorsal guard hair, etc. No attempt should ever be made to draw conclusions about pubic hairs based on observations of head hairs, for example.

Comparison process

Samples received for examination are initially inspected without the aid of any microscope, simply by viewing how they rest on a surface or holding them up to the light. They are next examined with a stereomicroscope noting color in reflected light, length, curliness, or anything of apparent significance. Usually at that time, hairs that appear to be representative of each sample as a whole or represent specific hairs for comparison are selected for additional study and mounted for microscopical analysis by polarized light microscopy, fluorescence microscopy and possibly by scanning electron microscopy or infrared microspectroscopy of adherent residues.

If the need for a detailed comparison is apparent, hairs from each exemplar are examined one at a time and thoroughly described with regard to all of the observable characteristics. The variation shown along each hair shaft is noted individually for each characteristic, and the process is repeated for each hair examined. A list of characteristics and traits can be used to assist with a systematic recording of features and their variations (**Tables 3–5**; **Figs 1–11**). A description of hair from each individual which expresses the range of traits observed for each characteristic and any significant special features or uncommon traits can be recorded.

The questioned hairs are likewise scrutinized and characterized microscopically. Consideration must be given to the origin of the evidence hair in deciding how to treat these samples. If several hairs were collected together from a source that makes it clear that they share a common origin, they may be treated as one sample and evaluated as a group in the same manner as the exemplars. If individual evidence hairs were collected from different sources, there is no justification for an assumption that they share a common origin. They must, therefore, be treated as several discreet samples, perhaps limited to a single hair each.

Finally, the selected hairs from the exemplar are compared with each questioned hair using a comparison microscope. The side-by-side analysis will either confirm or disprove the apparent similarities between known and questioned hair. The direct comparison is also a time to search for areas that show the same traits at corresponding points along the length of the hair shaft.

In other words, when the initial examination indicates a similar range of values for each characteristic between the known and questioned hair, it must now be determined whether the same individual traits for each characteristic can be juxtaposed at corresponding points along the questioned and known hair. If two samples of hair have a similar range of like features demonstrated during the initial separate microscopical comparisons using a stereomicroscope and polarizing microscope, and selected hairs are found to be alike for all or nearly all characteristics and traits in the side-by-side comparison using a comparison microscope, then a conclusion that the hairs are similar in all respects (a match) is justified and the samples could have shared a common origin.

Analysis

When comparing hairs, all characteristics of both the known and unknown specimens must be considered. A single significant difference between the two is a strong indication of two sources. Several repeated, fundamental dissimilarities establish with certainty that two specimens are not from a single individual.

Individualization rests therefore not only on a similar combination of identifying traits, a condition which always must be fulfilled, but also on a coexistent lack of any dissimilarities between the questioned and known hairs.

No two specimens of hair, even if from the same spot on the body, are identical in every detail since variation is an integral part of natural growth. The amount and kind of variation differs. With some it is slight and occurs only in subtle details; with others it covers a rather wide range. Variation does not preclude identification of the hair. In fact, variation around the basic qualities of the hair forms an additional factor, which serves to individualize hair.

Each characteristic may vary in any or all of the following ways. They may show variation at different points along a single hair shaft, between different hairs of the same individual, or between the hairs of different individuals. Hair of two different individuals may be entirely dissimilar or they may be very much alike but not identical. Individualities, many of them inconspicuous details, distinguish hair samples that appear very much alike. Everyone seems to appreciate that two specimens are not from the same individual when there are a vast number of differences, but only a few fundamental dissimilarities may not seem to lead to so positive a conclusion; nevertheless, they do. If two specimens are from a

Table 3 Human hair characteristics and traits

Feature	*Characteristic*	*Trait*
Color	Hue	Colorless
		Blond
		Red
		Auburn brown
		Golden brown
		Ashen brown
		Black
	Pigment density	Absent
		Light
		Medium
		Heavy
		Opaque
	Pigment distribution	Uniform
		Peripheral
		One-sided
		Random
		Central/medial
		Cuticle pigment
		Banded
	Pigment aggregation	Streaked
		Clumped
		Patchy
	Pigment aggregate size	Large
		Medium
		Small
	Pigment shape	Round
		Oval
		Oblong
	Pigment size	Large
		Medium
		Small
	Artificial treatment	Dyed
		Bleached
		Other
Structure	Length	
	Diameter	
	Curvature	
	Form	Straight
		Wavy
		Curly
		Twisted
		Tightly coiled
		Crimped
	Cross-sectional shape	Round
		Oval
		Triangular
		Flattened
		Pear-shaped
	Shaft configurations	Buckling
		Convoluting
		Shouldering
		Undulating
		Splitting
	Cuticular thickness	Thin
		Medium
		Thick
	Outer cuticle margin	Smooth
		Flattened
		Serrated
		Cracked
		Looped
	Inner cuticle margin	Distinct
		Smooth
		Cracked
		Indistinct
	Cortical texture	Coarse
		Uniform
	Ovoid bodies	Size
		Shape
		Color
		Distribution
		Abundance
	Cortical fusi	Size
		Shape
		Distribution
		Abundance
	Medulla	Continuous
		Interrupted
		Fragmental
		Opaque
		Translucent
		Relative width
		Amorphous
		Patterned
		Absent
Treatment	Root present	Telogen
		Catagen
		Anagen
		Sheathed
		Follicular tag
		Post-mortem band
	Root absent	Cut
		Broken
		Decomposed
		Crushed
	Distal ends	Tapered
		Rounded
		Abraded
		Square
		Angle
		Frayed
		Split
		Crushed
		Broken
		Cut
		Singed
	Alterations	Straightening
		Waving
		Curling
	Artifacts	Lice
		Mold/fungus
		Bite marks
		Debris
		Blood
	Abnormalities	Pili annulati
		Trichoschisis
		Monilethrix
		Trichorrhexis nodosa
		Trichorrhexis invaginati
		Pili torti
		Trichonodosis
		Cartilage hair hypoplasia

Table 4 Animal hair characteristics and traits

Feature	*Characteristic*	*Trait*
Color	Color	Bicolored
		Multicolored
		Banded
		Unbanded
	Color band configuration	Single (one light band)
		Triple (two light and one dark band)
		5-banded (three light and two dark bands)
	Color band location	Upper shield
		Central shield
		Lower shield
		Lower shaft
		Upper shaft
Shaft	Undulations	
	Basal configuration	Wine-glass shaped
		Slightly tapered
	Length	
	Diameter	
	Spatulate shield	In upper shaft
		In middle shaft
	Flattened shaft	In lower shaft
	Stricture	Lower shaft
		Subshield
	Cross-sectional shape and appearance	Circular medium medulla
		Circular large medulla
		Oval large medulla
		Oval medium medulla
		Oval medulla absent
		Eye-shaped
		Oblong large medulla
		Oblong medium medulla
		Cigar-shaped
		Concavoconvex divided medulla
		Concavoconvex bilobed medulla
		Concavoconvex large medulla
		Reniform
		Dumbbell shaped
Medulla	Medulla	Opaque
		Infiltrated
		Width
		Medullar index
	Medulla type	Absent
		Simple unbroken amorphous
		Unbroken cellular
		Unbroken vacuolated
		Unbroken with cortical intrusions
		Unbroken lattice
		Narrow medulla lattice
		Wide medulla lattice
		Narrow aeriform lattice
		Wide aeriform lattice
		Uniserial ladder
		Multiserial ladder
		Interrupted
		Fragmental
		Globular
		Stellate
		Intruding
Scales	Scale pattern	Simple coronal
		Regular petal
		Broad petal
		Diamond petal

Table 4 continued

Feature	*Characteristic*	*Trait*
		Narrow diamond petal
		Irregular petal
		Single chevron
		Double chevron
		Regular mosaic
		Irregular waved mosaic
		Flattened irregular mosaic
		Regular wave
		Irregular wave
		Streaked wave
		Pectinate
	Shape	Spatulate
		Flattened and tapered
		Round to oval
	Scale margin	Smooth
		Crenate
		Rippled
		Scalloped
		Dentate

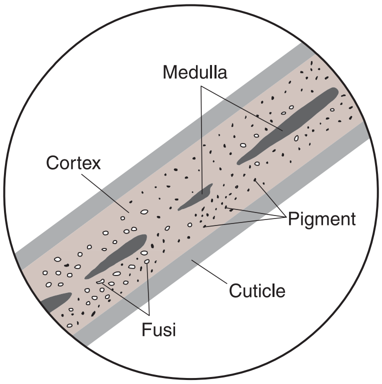


Figure 1 Hair shaft. The hair shaft of human hair as viewed microscopically in a mountant. All hairs, whether human or domesticated or wild animal, have the same structure which includes a cuticle, cortex, pigment, fusi, and medulla.

single source, then no fundamental differences should exist. Conversely, if there is any basic dissimilarity which cannot be accounted for, then the two specimens must have been obtained from different sources. Hair comparison involves the discovery and study of all identifying characteristics; the differentiation between those which are typical and those which are abnormal or represent the unusual; recognition of those which are disguised or have been deliberately and consciously changed; and determination of the normal amount of variation common to the particular sample. See **Table 6** for summary of fundamental concepts of a microscopical hair comparison.

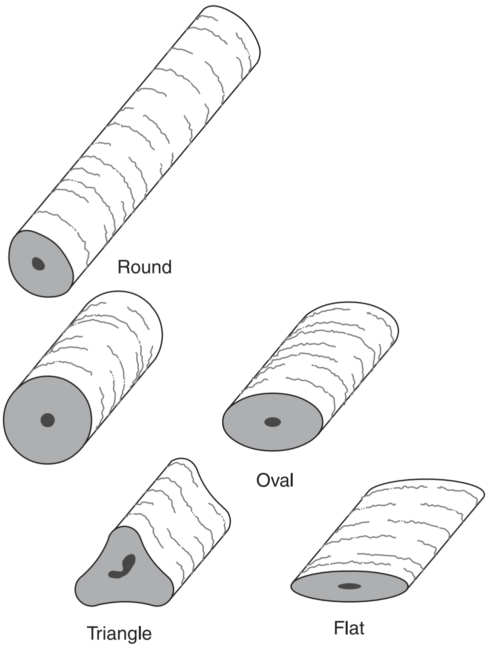


Figure 2 Cross-sectional shapes. Microscopically, hairs often appear as two-dimensional objects, but they all have a third dimension represented by a cross-sectional shape. Human hairs range in shape from round to flat. Beard hairs are often triangular in cross section.

Quality Assurance

When a questioned hair is said to be consistent with a known source, there are two possibilities. Either the hair actually originated from that source or there was a coincidental match. Since it is possible for two

Table 5 Terminology for microscopic hair comparisons

Term	*Definition*
Central region	Toward the core area of the hair shaft, near where the medulla is found
Characteristic	Any feature which may be useful for identification and/or comparison
Comparison	The process of examining two or more hairs for the purpose of either identifying them as having come from the same class of hairs or attempting to associate them with or disassociate them from a given individual
Competency test	Tests given at the end of the training period, successful completion of which demonstrates an examiner's ability to perform forensic hair examinations in casework
Dissimilar	Significant unexplained differences between questioned and known hairs
Distal end	The end of the hair distant from the root
Feature	Any morphological part of a hair that reflects the hair's appearance, texture, or microscopic characteristics and traits
Identification	The process of determining that a given hair belongs to or came from a defined class of hairs (e.g. species, somatic or racial origin)
Individualization	The process of determining that a given hair came from one particular (individual) source to the exclusion of all other similar sources
Known	A sample taken as representative of a particular body area of a specific individual
Limited sample	A sample of hairs known not to be of sufficient number or quality to represent all possible characteristics or traits
Match	The association of known and questioned hairs by showing that the hairs are similar in all respects
Medial region	The portion of the hair intermediate between the proximal and distal ends
Peer review	A confirmation process where a technical peer (colleague) independently reviews the conclusions and supporting data from the hair examiner before it is reported
Peripheral region	Toward the outermost areas of the hair, including the cuticle and the cortex distant from the medulla and near the cuticle
Practical examination	Practice tests given during the training period
Proficiency test	A quality assurance measure used to monitor performance and identify areas where improvement may be needed
Proximal end	The end of the hair nearest the root
Questioned	A sample of unknown origin collected for the purpose of identification and/or comparison with a known sample
Range	The complete set of values or traits exhibited by the hair of one individual with regard to a specific characteristic
Representative sample	A sample of hairs that is expected, from a statistical standpoint, to represent the range of values or traits
Sample	One or more hairs used for identification, comparison or reference
Similar	A conclusion when the combination of microscopic characteristics, values, and traits are exhibited by the known and the questioned hair sample with no significant unexplained differences; or when the known and questioned hairs are microscopically indistinguishable
Trait/value	The qualitative or quantitative assessment of a particular variable characteristic based on a single observation
Verification	A confirmation process where a second qualified hair examiner microscopically reviews the results from a comparison, employing the same methodology as the first examiner, and is free to agree or disagree with the conclusions of the first

different individuals to have hairs that are microscopically indistinguishable, it is known that coincidental matches can occur in forensic hair comparisons.

One way to increase reliability and limit the number of coincidental matches is to insure the hair examiner is: properly and thoroughly trained, experienced, allowed time to reach the decision, allowed negative and inconclusive determinations when evidence available does not warrant more, allowed access to the background and investigative information, and allowed to consult with a colleague to verify the analysis. A better quality assurance technique is verification of conclusions by another qualified examiner. Peer review is less valuable than the redundant analysis required by verification. Proficiency tests, which are difficult to design, are useful for training, practice and gaining a better understanding of the comparison process, but they cannot replace peer review and verification.

Where possible, laboratories should employ a confirmation or verification process whereby the hair examiner, having determined that questioned hairs from items of evidence and known hair standards exhibit the same microscopic characteristics, then takes these hairs to another qualified examiner. The second hair examiner reviews the hairs micro-

Table 6 Fundamental concepts of microscopic hair comparisons

Concept	*Definition*
1	All observable characteristics of both the known and unknown hair specimens must be considered
2	A comparison involves the search for basic dissimilarities which cannot be accounted for by a logical, common sense explanation
3	A single significant difference between the two is a strong indication of two sources
4	Several repeated fundamental dissimilarities establish without a doubt that two specimens are not from a single individual
5	Sometimes inconspicuous traits distinguish the hair of two samples that appear very much alike otherwise
6	No two specimens of hair from one individual are identical in every detail
7	Known and questioned hairs need not be identical in the sense that the two sets of hair can be matched bit by bit
8	Differences between the questioned and known hairs cannot exceed the variation found in the known sample
9	As with all types of associative evidence, quality exemplars are mandatory and no conclusions are possible without them
10	Under no circumstances can an association be established by one, two, or even several unusual characteristics alone
11	A combination of a sufficient number of points of agreement with a coexistent lack of basic divergencies between the questioned and known hair is required for an association
12	If the questioned hairs were mixed with the exemplar, and the hairs could not be distinguished or retrieved, the association is justified
13	The degree of certainty given an association depends on the individuality of the hair and the amount of questioned hair available for comparison
14	The conclusions drawn from a forensic hair comparison are based on statistical considerations, related experience, scientific taste, and forensic judgment
15	The conclusions must be reasonable because their impact is determined by their convincingness
16	All associations are accepted provisionally. The provisional associations are tested by examining alibi and elimination samples. When these alibi samples are scientifically eliminated, the association gains probative value

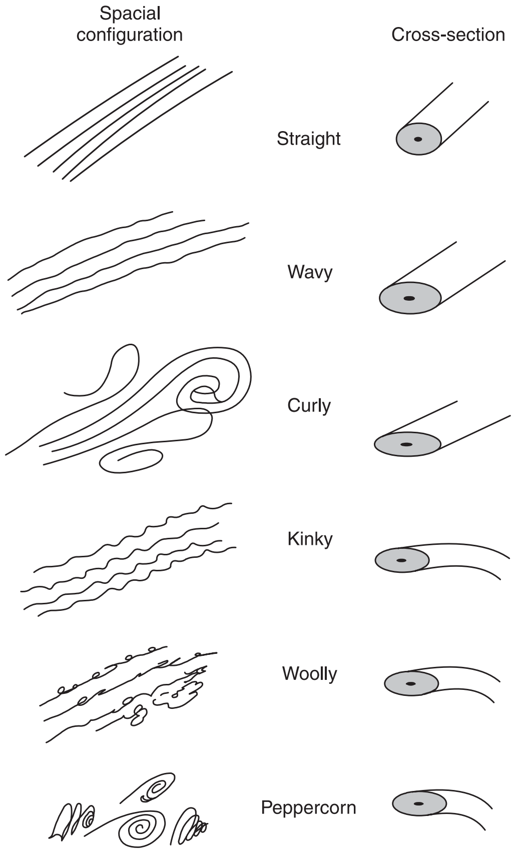


Figure 3 Human hair forms. The cross-sectional shape largely determines the spatial configuration. Hair forms for different human populations vary from straight with a round cross-section to peppercorn with a flat twisting cross-section.

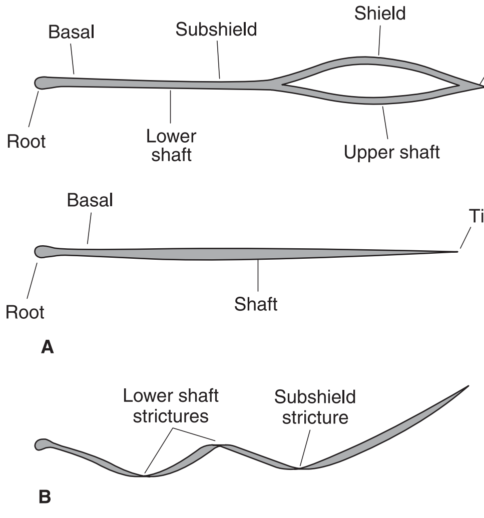


Figure 4 Regions of shielded and nonshielded hair. The parts (**A**) and locations of strictures (**B**) of a common animal hair shaft are noted from tip to root including the blade.

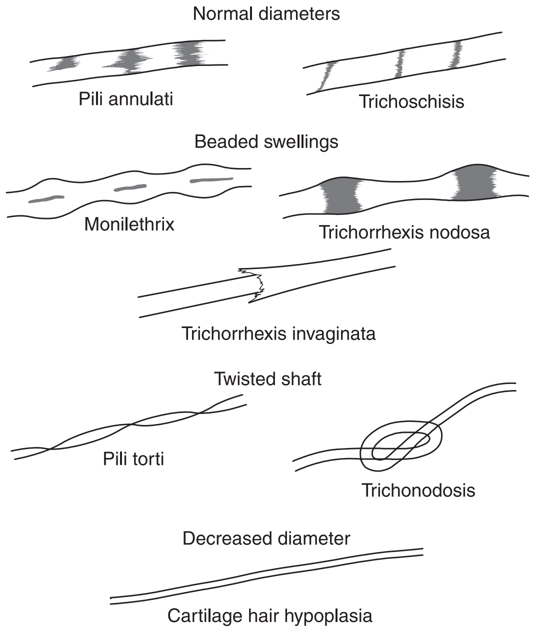


Figure 5 Hair shaft abnormalities. Sometimes human hair grows in abnormal ways due to disease. These hairs, when viewed microscopically, sometimes have normal diameters with crossway markings, beaded swellings, twisted shafts or decreased diameters.

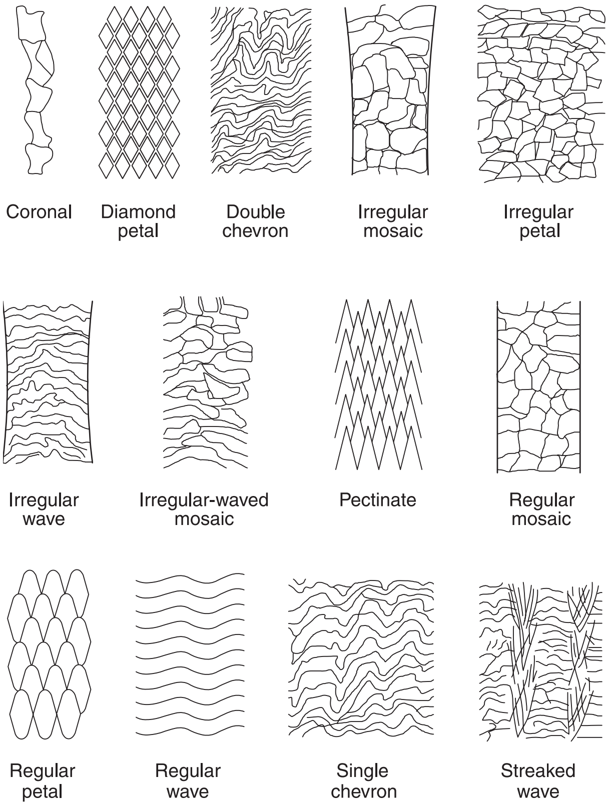


Figure 7 Scale patterns. Details of patterns seen on hairs of different families of animals.

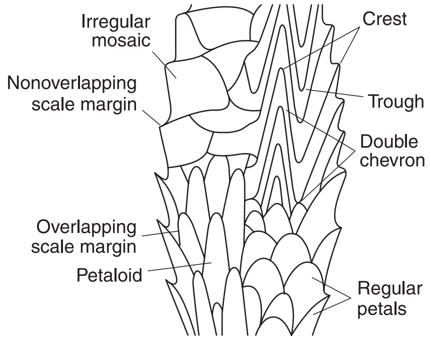


Figure 6 Scale terminology and patterns. The scale patterns on the hair of wild animals vary greatly between different families of animals and sometimes along the length of a single hair. These patterns can be visualized when the surface scales are reproduced with a cast or viewed with a scanning electron microscope. (Redrawn with permission of the State of Wyoming from Moore TD, Spence LE and Dugnolle CE (1974) *Identification of the Dorsal Guard Hairs of Some Mammals of Wyoming*. Cheyenne, WY: Wyoming Game and Fish Department.)

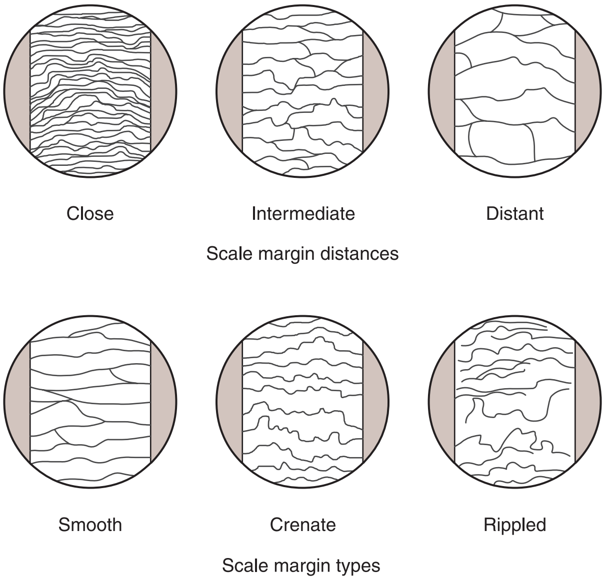


Figure 8 Scale margins. At higher magnifications, three common patterns on the leading edge (exposed margins) of animal hair scales are found. The distances between the scale margins can be described as close, intermediate or distant or simply measured with a microscope.

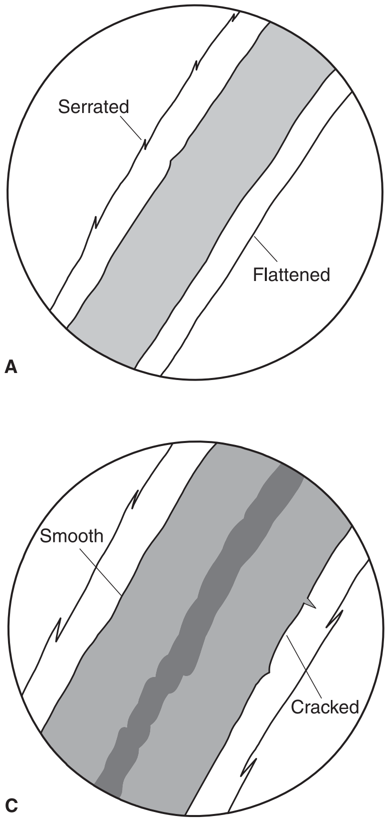


Figure 9 Outer and inner cuticular margins. When viewed with transmitted light, in a medium with a refractive index near that of hair, the cuticle of human hairs appears as a transparent thin band on either side of the hair shaft. The outer cuticular margins (outside surface of the hair shaft) (**A, B**) can be flattened, serrated, cracked or ragged whereas the inner margin (interface between cuticle and cortex) (**C**), if distinct, can be smooth or cracked.

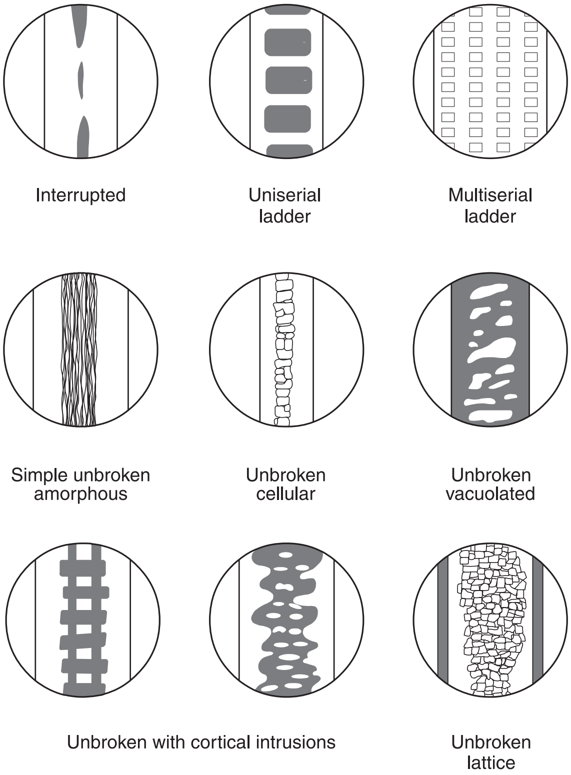


Figure 10 Medulla types. The medulla in animals varies greatly. In humans the medulla is always less than half the width of the hair shaft; in animals it usually fills more than half of the shaft diameter. The various patterns for medullae found in domesticated and wild animals are shown.

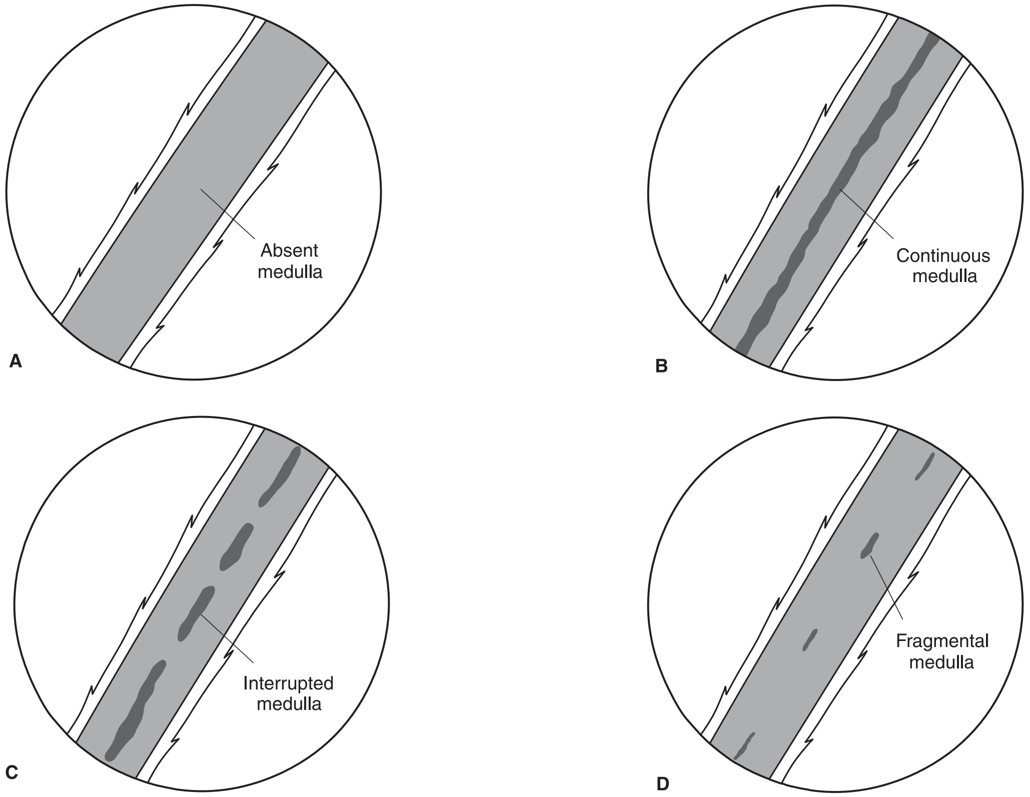


Figure 11 Medullation. In humans the medulla may be absent (**A**), continuous (**B**) or discontinuous (**C, D**). In an interrupted, medulla (**C**) the length of the medullary band is greater than the length of the space between. In a fragmental medulla (**D**) the medullary band length is shorter than the distance between.

scopically, employing the same methodology as the first hair examiner. The second examiner is free to agree or disagree with the first examiner's results, but only when there is agreement are the hairs associated.

Conclusions

In order to conclude that a questioned hair is associated with a known source, it must first be determined that the characteristics exhibited by the questioned hair fit within the range of characteristics present in the known sample. In addition, an association can be strengthened if there is also a one to one correspondence of all characteristics between the questioned hair and one or two known hairs with no significant differences over the entire length of the hair.

See also: **Analytical Techniques**: Microscopy; Spectroscopy: Basic Principles. **Hair**: Overview; Identification of Human and Animal Hair; Comparison: Other; Comparison: Significance of Hair Evidence; Deoxyribonucleic Acid Typing.

Further Reading

Appleyard HM (1978) *Guide to the Identification of Animal Fibers*. Leeds: Wool Industries Research Association.

Bisbing RE (1982) The forensic identification of human hair. In: Saferstein R (ed.) *Forensic Science Handbook*, pp. 184–221. Englewood Cliffs, NJ: Prentice-Hall.

Brunner H and Coman BJ (1974) *The Identification of Mammalian Hair*. Melbourne: Inkata Press.

Chayko GM, Gulliver ED and Macdougall DV (1991) Trace evidence. In: *Forensic Evidence in Canada*, pp. 259–299. Aurora, Ontario: Canada Law Book.

DeForest PR (1982) Foundations of forensic microscopy. In: Saferstein R (ed.) *Forensic Science Handbook*, pp. 416–528. Englewood Cliffs, NJ: Prentice-Hall.

Hicks JW (1977) *Microscopy of Hairs: A Practical Guide and Manual*. Washington, DC: US Government Printing Office.

Kirk PL (1974) In: Thornton JI (ed.) *Crime Investigation*, 2nd edn. New York: Wiley.

Kobori T and Montagna W (1976) *Biology and Disease of the Hair*. Baltimore: University Park Press.

Lee HC and DeForest PR (1984) Forensic hair examination. In: *Forensic Sciences*, 37A: pp. 1–46. New York:

Matthew Bender.
McCrone WC and Delly JG (1973) *The Particle Atlas*, 2nd edn. Ann Arbor: Ann Arbor Science Publishers.
Moenessens AA, Inbau FE and Starrs JE (1986) *Scientific Evidence in Criminal Cases*, 3rd edn. Mineola, NY: The Foundation Press.
Montagna W and Parakkal PF (1974) *The Structure and Function of Skin*. New York: Academic Press.
Montagna W and Dobson RL (1969) *Hair Growth*. London: Pergamon Press.
Moore TD, Spence LE and Dugnolle CE (1974) *Identification of the Dorsal Guard Hairs of Some Mammals of Wyoming*. Cheyenne, WY: Wyoming Game and Fish Department.
Robbins CR (1979) *Chemical and Physical Behavior of Human Hair*. New York: Van Nostrand Reinhold.
Savill A and Warren C (1962) *The Hair and Scalp*, 5th edn. Baltimore: Williams and Wilkins.
Wildman AB (1954) *The Microscopy of Animal Textile Fibres*. Torridon, UK: Wool Industries Research Association.

Comparison: Other

B D Gaudette, Royal Canadian Mounted Police, Ottawa, Canada

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0561

Introduction

The most reliable and widely used methods of forensic hair comparison involve microscopic analysis and, more recently, DNA analysis. However, a wide variety of other methods have been proposed and find use in special applications, usually as an augmentation of conventional methods. These techniques can be divided into four broad and somewhat overlapping categories: microchemical methods, instrumental chemical methods, physical methods and biochemical methods.

Microchemical methods

The reverse of detecting treatments on hairs is to deliberately treat a hair with various dyes or reagents and then microscopically observe and compare the changes produced. In one example of such a microchemical testing technique, changes produced on hairs by treatment with mercaptoacetic acid were observed by scanning electron microscopy. Another example used Lucifer Yellow CH dye. These examples suggest that further research into this technique would be useful.

Instrumental Chemical Methods

Many instrumental chemical methods can play a role in forensic hair comparison. The most promising of these are the methods used to examine cosmetic treatments on hairs. Other instrumental chemical methods are targeted at naturally occurring or environmentally deposited components of the hair.

The most widely reported of this latter group of methods is pyrolysis gas chromatography (Py-GC). Although early work with packed column Py-GC failed to provide useful results, it was felt that the enhanced resolving power of capillary column Py-GC would provide a potential method of individualizing human hair. It was found, however, that the relative amounts of the pyrolysis products of hair proteins were the same from person to person, and that the same sort of variability occurred within a set of samples from the same individual as within a set of samples from different individuals. A subsequent study, found three major components (benzene, toluene, and styrene) whose capillary column Py-GC pyrograms differed significantly between individuals and remained constant over time. In a blind trial hairs could be correctly sorted into groups and it was concluded that although the method was not suitable for individualizing hairs on its own, it could prove useful in combination with other methods.

A promising application of Py-GC is the analysis of nicotine in hair. In a study of 48 subjects an increasing nicotine concentration gradient was found from root to tip, indicating adsorption of nicotine from the outside. A successful distinction was made between smokers and nonsmokers and a good correlation was shown between nicotine concentration and self-reported exposure. This method has potential to provide useful information to aid investigations (smoking status) as well as additional comparison characteristics.

Many different instrumental methods are currently used to analyze trace elements in human hair. Although this is a viable method for environmental and medical studies, there are many problems associated with its use in forensic hair comparison. Trace elements in human hair can arise from metabolism, cosmetic preparations or the external environment. Trace element content of hair is affected by personal factors (age, sex, diet, etc.), external factors (shampoos, air-borne contaminants, etc.), frequency of hair washing, the structure of the hair itself, and by sample treatment prior to analysis. Furthermore, hair trace

element content varies with location on the body, from hair to hair on the same body, from root to tip along the hair shaft, across the diameter of the hair, and according to the stage in the hair growth cycle. Because of such variation and sample treatment problems, trace element analysis of hair is not considered to be a reliable method of forensic hair comparison.

A major use of chemical instrumental methods has been drug profiling of human hair. Reports of opiates, barbiturates, methaqualone, codeine, cocaine, morphine and phencyclidine being detected in hair have appeared in the literature in recent years. Radioimmunoassay has been the most common method of analysis. All work to date has concentrated on the toxicological use of hair analysis for drugs. The large number of licit and illicit drugs available and the correlation between crimes of violence and drug abuse make this a promising area for further research as a method for comparing known and questioned hair samples.

The use of GC-MS to detect oxidative dyes and other hair products was discussed elsewhere. Fourier transform infrared spectroscopy (FTIR) also has potential in the analysis of hair sprays and other cosmetic treatments. High pressure liquid chromatography (HPLC) may also prove useful for the analysis of hair dyes and rinses.

Much current work in instrumental analysis is related to applications of combined instruments such as gas chromatography/mass spectrometry (GC-MS) and HPLC-FTIR. It is possible that the future will bring applications of these combined instrumental methods to forensic hair comparison.

Other instrumental chemical methods have been reported in single studies which have not been followed-up or generally adopted.

Physical Methods

The use of physical properties of hair in forensic hair comparison has been limited. The potential of spectrofluorometry in the comparison of photoluminescent properties was pointed out in 1975 and the significance of some elastic constants in hair comparison has been investigated. Many years ago, density was proposed as a hair comparison characteristic. Those physical properties thus far investigated have been found to have large within-hair and within-individual variations which have tended to discourage their use. The sophisticated computers and statistical packages of today and the future might be able to overcome this problem, provided that these variations are significantly less than the variations that exist between individuals.

Biochemical Methods

Biochemical methods provide some promising areas for future research in forensic hair comparison. For several years, ABO grouping of hairs has been successfully used routinely in Japan and other East Asian countries. However, when applied to non-Mongoloid hairs, ABO grouping has been found to give results that are somewhat erratic. Japanese workers have shown that blood group substances appear to be localized in the medulla. This knowledge could lead to future research which might increase the reliability of ABO grouping with Caucasian and Negroid hairs.

For several years, enzyme typing of hair root sheaths was a method routinely used in many forensic laboratories. Phosphoglucomutase, esterase D and glyoxylase were the systems most commonly used. Other workers developed a procedure for simultaneous typing of erythrocyte acid phosphatase, adenylate kinase and adenosine deaminase in human hair root sheaths. However, since root sheaths occur in only a small percentage of questioned hair, the application of this method is somewhat limited. As with blood and semen analysis, in recent years DNA-based methods have superseded enzyme typing methods for hair.

Isoelectric focusing in polyacrylamide gels for non-carboxymethylated keratins has proven useful in species determination and could be used to detect individual variation in a given population. It has been suggested that sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) could be applied to forensic hair comparison. It was noted, however, that further research was needed. Following a good discussion of hair comparison by electrophoresis another worker concluded that: 'The results of a study of a limited number of human hair samples by two-dimensional electrophoresis are sufficiently encouraging to recommend that a trial be started in which hair samples from a much larger number of individuals be examined ... Further research is required to assess the potential of the technique but present indications are that it may become a useful supplement to microscopic description of hair.' (Marshall, 1984)

Amino acids analysis is another biochemical method with some future possibilities, although the variability in amino acid composition of hairs may cause problems.

Concluding Remarks

Let us now look at the effect some of the previously discussed other analytical methods might have on evidential value. Making as many sets of independent

comparisons as possible will greatly reduce the probability of coincidental matches. By doing so, provided they do not also greatly increase the probability of examiner errors, these other analytical methodologies should lead to a large increase in the evidential value of forensic hair comparison.

Two notes of caution should, however, be introduced when considering any future research on hair comparison methodology. First, the probabilities of type I errors (incorrect elimination) and type II errors (incorrect association) vary inversely. As the probability of type II errors is decreased by new methods, the probability of type I errors will often increase. Accordingly, in evaluating new methods we need to ask whether or not the method significantly decreases the probability of type II errors and, if so, if it also increases the probability of type I errors to an unreasonable level.

A second reservation or caution must be expressed with regard to research on hair characteristics which are sensitive to the environment or consumer products. If, for example, a suspect is not apprehended shortly after a crime, his or her hair could have been subject to cosmetic treatment between commission of the crime and the subsequent submission of the known hair sample to a forensic laboratory. Thus the cosmetic treatment characteristics of the known hair sample could be either deliberately or accidentally altered. When compared with a questioned hair found at the crime scene or in the victim's clothing, a type I error would likely result if too much emphasis were placed on cosmetic treatment characteristics in such circumstances.

Collaborative studies provide an excellent means of assessing the costs and benefits of any new hair comparison methodology.

See also: **Analytical Techniques**: Mass Spectrometry. **Drugs of Abuse**: Hair. **Hair**: Overview; Hair Transfer, Persistence and Recovery; Identification of Human and Animal Hair; Comparison: Microscopic; Comparison: Other; Comparison: Significance of Hair Evidence; Deoxyribonucleic Acid Typing; **Microchemistry**.

Further Reading

Baumgartner W, Hill V and Blahd W (1989) Hair analysis for drugs of abuse. *Journal of Forensic Sciences* **34**:1433–1453.

Choudhry M, Kingston C, Kobilinsky L and DeForest P (1983) Individual characteristics in chemically modified hairs revealed by scanning electron microscopy. *Journal of Forensic Sciences* **28**:293–306.

Goin L, McKee W and Kirk P (1952) Application of the microdetermination of comparative density. *Journal of Criminal Law and Criminology* **43**:263–273.

Ishuazawa F and Misawa S (1990) Capillary column pyrolysis gas chromatography of hair: a short study in personal identification. *Journal of the Forensic Science Society* **30**:201–209.

Johri M and Jatar D (1981) Significance of some elastic constants in the individualization of human head hair. *Police Research and Development* **3**:22–26.

Jones P (1975) New applications of photoluminescence techniques for forensic science. *Forensic Science A.C.S. Symposium Series 13*, pp. 183–184 and 191–196. Washington DC: American Chemical Society.

Kintz P (1996) *Drug Testing in Hair*. Boca Raton: CRC Press.

Loxley D (1984) The use of Lucifer yellow CH dye to stain human hair scales. *International Association of Forensic Sciences 10th Trienial Meeting*, Oxford.

Marshall R (1984) Forensic identification of hairs by electrophoresis. *International Association of Forensic Sciences 10th Trienial Meeting*, Oxford.

Munson T (1987) The status of hair comparisons by pyrolysis gas chromatography. *Crime Laboratory Digest* **14**:153–154.

Rodrigues-Calvo M, Carracedo A, Munoz I and Concheiro L (1992) Isoelectric focusing of human hair keratins: correlation with sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) patterns and effects of cosmetic treatments. *Journal of Forensic Sciences* **37**:425–431.

Zahlesen K and Nilsen O (1990) Gas chromatographic analysis of nicotine in hair. *Environmental Technology* **11**:353–364.

Comparison: Significance of Hair Evidence

B D Gaudette, Royal Canadian Mounted Police, Ottawa, Canada

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0562

Introduction

Forensic hair examinations are most frequently conducted to assist investigation and prosecution of crimes of violence such as murders and sexual assaults. Although the probative value of hair comparison evidence is generally much lower than that of some other forms of forensic evidence such as fingerprints or DNA, it can still provide good corroborative evidence. Hair comparison can help establish associations between any combination of the following: accused, victim, crime scene or weapon.

Unlike blood, the mere presence of hair on an exhibit item is not usually by itself of any evidential value. A hair comparison with a sample of known origin must first be performed. After making a forensic hair comparison, the next (and most important) step is to assess its significance. This involves a two-stage process: using the results to develop a conclusion, and then interpreting that conclusion to form an expert opinion as expressed in a report or court testimony. This article begins with a general discussion on evaluating associative forensic science evidence. A discussion is then given of studies on probabilities and human hair comparison and the criticisms that have been made of them. Some other studies of the value of hair comparison evidence are then discussed followed by a discussion of report writing and court testimony.

Evaluating Associative Forensic Evidence

The fundamental question to consider when evaluating associative forensic science evidence is 'What is the value of the evidence in establishing a particular association?' This question is not as simple and straightforward as it might first appear. First, there is considerable controversy as to whether or not it is the role of a forensic scientist to answer this question in a court setting. However, even those who feel that it is not should acknowledge that consideration of such concepts helps to clarify thinking outside the court environment. Second, there are those who persuasively argue that evidential value can only be determined through Bayesian analysis, whereas others seem to find such analysis difficult to understand or accept. Finally, there are many components to the fundamental question such as:

1. What is the probability that the association was due to coincidence?
2. What is the probability that the association was due to examiner error?
3. What is the probability that there is an alternative explanation for the evidence such as secondary transfer, contamination or deliberate planting?

Discussion of the above noted controversies is beyond the scope of this article. At this point, it will simply be noted that from a Bayesian perspective, the important questions are: (a) what is the probability of the hair evidence if there was association; and (b) what is the probability of the hair evidence if there was no association, and that it is the ratio of these two probabilities that determines evidential value.

The concept of type I and type II errors can also be helpful in understanding evidential value. Note that there are two possible states of nature with regard to association. Either there was some form of association (denoted as A) or there was not (denoted as N). Ignoring inconclusive results, there are two possible outcomes of a forensic scientist's examination: either the evidence indicated association (E) or it did not ($\bar{E}$). If the state of nature is A and the forensic scientist gives an opinion indicating E, the forensic scientist is correct. Similarly, when the state of nature is N and the forensic scientist says $\bar{E}$, he or she is also correct. However, if the state of nature is A and the forensic scientist says $\bar{E}$, a type I error or incorrect exclusion has occurred. If the state of nature is N and the forensic scientist has given an opinion indicating $\bar{E}$, there was a type II error or incorrect association.

Type I and type II errors can be better understood through an analogy to a fire alarm. Type I errors correspond to the fire alarm not ringing when there is a fire. Type II errors correspond to the fire alarm ringing when there is no fire. Depending on the decision we are making and the constraints involved, knowledge of the probability of type I or type II errors, or both, can greatly assist in making the decision. With the fire alarm, a type I error would be more serious than a type II error; with hair comparison a type II error would be the more serious since it could result in wrongly incriminating evidence being presented against a suspect. (It is for this reason that hair examiners should set a level of discrimination that minimizes type II errors without incurring an unreasonable number of type I errors.)

In report writing and court testimony, once a questioned hair has been found to be consistent with a known sample, it is only the probability of type II errors that is important in evaluating the evidence, just as once a fire alarm has rung it is the probability of type II errors and not the probability of type I errors that influences our decision as to whether or not to leave the building. In attempting to determine the value of forensic hair comparison in establishing associations, we need to know the probability of type II errors due to coincidental matches, examiner errors and alternative explanations for the evidence.

Studies on Probabilities and Human Hair Comparison

A lay person might observe that there are quite a few differences in the gross appearance of people's hair. Some people have long hair, some short, some have curly hair, some straight; some have dark-colored hair, some light; some people bleach and dye their hair, some do not. A lay person would not, however, have any idea of the intrapersonal variation in hairs

and the large number of hair characteristics that can be observed microscopically and the number of variables each characteristic can have. Accordingly, a lay person would not have any intuitive feel for the average value of forensic hair comparison evidence. In an attempt to rectify this situation, Gaudette and Keeping conducted a study in which, with the aid of a card-coding system, 366 630 pairwise comparisons were made between 861 hairs from 100 individuals. Of these, nine pairs of hairs were found to be macroscopically and microscopically indistinguishable. From this it was calculated that if a single scalp hair selected at random from individual A was found to be consistent with a single hair selected at random from individual B, the chance that the match was due to coincidence was about 9/366 630 or 1/40 500. If a single hair selected at random from A was found to be consistent with a representative known sample from B (consisting in the study of an average of about nine mutually dissimilar hairs), on average the chance of a coincidental match was 9 × 1/40 500 or about 1 in 4500.

In a similar study with pubic hairs, 101 368 comparisons were made of 454 hairs from 60 individuals. It was found that 16 pairs of hairs were macroscopically and microscopically indistinguishable. Therefore, if a single pubic hair selected at random from person A was found to be consistent with a single pubic hair selected at random from individual B, an estimate of the average probability of a coincidental match would be about 16/100 368 or 1/6336. If the single hair selected at random from A was found to be consistent with a known sample of pubic hairs (which in the study consisted of about eight mutually dissimilar hairs) from B, an estimate of the average probability of that one hair having originated from someone else would be 8 × 1/6336 or about 1 in 800. The greater likelihood of a coincidental match for pubic hair than for scalp hair may reflect the smaller variation in characteristics of pubic hairs throughout the population.

An interesting finding of the pubic hair study was that hairs from one individual were involved in three matching pairs of hairs while hairs from seven other individuals were involved in two matching pairs. This shows that certain hair types and certain individuals are more likely to be involved in coincidental hair matches than others.

The Gaudette and Keeping results refer to the situation where a single questioned hair is found to be consistent with a known sample. The finding of two or more questioned hairs to be consistent with the known sample will greatly reduce the probability of a coincidental match. A probability estimate cannot be obtained by simply multiplying 1 in 4500 by 1 in 4500, however, since independence cannot be assumed.

It should be emphasized that the Gaudette and Keeping probability results are average values made up of the sum total of all hair types – from unusual hairs (where probability of a coincidental match would be virtually 0) to hairs of average commonness (where the probability of a coincidental match would approximate 1/4500), to common featureless hairs (whose probability of a coincidental match would be considerably greater).

The Gaudette and Keeping results can provide a good estimate of the average value of hair comparison evidence in establishing associations when the following conditions are met.

1. The probability of examiner error is very low. (This condition should be met when a well-trained qualified examiner carefully conducts the examination.)
2. The probability of secondary transfer, contamination or deliberate planting of evidence is very low.
3. Caucasian hairs are involved.

Those using such probability estimates should take care, however. It is extremely important to word probability statements carefully.

The Gaudette and Keeping studies were criticized as it was claimed they contained defects in experimental design and improper statistical treatment of the data. These criticisms have been further rebutted. Although Gaudette and Keeping's work has been criticized, no studies have been offered to refute the results, and it has not been claimed that hair comparison evidence is not good evidence.

Other Studies on the Value of Hair Comparison Evidence

A wide range of opinions as to the value of hair evidence has appeared in the literature. Some authors take a disparaging view of hair evidence. The following quotation is typical: 'There is nothing about hair comparable to the specificity of fingerprints, and at best the probability of establishing identification from hair is perhaps no greater than the probability of determining identification using the ABO blood group system in blood smears.' (Camps 1968). On the other hand, the following quotation is typical of those authors who consider hair comparison evidence to have a high value: 'From research studies, it has been shown that hairs from two individuals are distinguishable and that no accidental or coincidental matches occurred, and would, therefore in actual casework be a relatively rare event.' (Strauss 1983).

The generally prevailing view of the value of hair comparison evidence lies between these extremes. These two quotations are representative:

> Through hair comparison it is presently only rarely possible to determine that a questioned hair did or did not originate from a particular person. In the vast majority of cases it can only be stated that a questioned hair is or is not consistent with having originated from a particular person. Accordingly, hair comparison evidence is generally only of value when used in conjunction with other evidence. (Gaudette 1985)

> 1) So far, a hair or hairs have not been shown to have any features exclusively confined to an individual; 2) Any indication of identity based on an examination of hair can therefore only be established in terms of probability; 3) The probability is increased, under certain circumstances, if all the characteristic elements are considered and is increased to an even greater extent when unusual features such as uncommon colours, disease, etc. are present. (Martin 1957)

Although a large number of individuals have expressed opinions as to the value of hair comparison evidence, actual research studies on the topic have been more limited. In addition to the work of Gaudette and Keeping, the following studies have been reported.

In 1940, Kirk reported that a group of his students were, without exception, able to match one questioned hair to the correct known sample in a group of 20, all of similar colour and from individuals of similar age.

In 1978, Gaudette discussed two additional experiments on the value of hair comparison evidence. In the first experiments, 100 randomly selected questioned hairs were compared in a blind trial to one known sample. This experiment was repeated three times with three trainees, each near the end of a one-year training period. Two of these trainees correctly chose the one and only hair that was consistent with the known sample. The third trainee first concluded that four of the questioned hairs were consistent with the known sample. After examining the hairs more closely and consulting with other examiners he was easily able to identify one of his choices as being incorrect, leaving three hairs he thought to be consistent with the known sample: the correct one and two others. When Gaudette examined the hairs, he stated that one of the two others could be eliminated but the remaining one was indistinguishable from hairs in the known sample. Another experienced examiner then studied the hairs and also concluded that one of the two others could be eliminated. This time, however, it was the one opposite to that picked by Gaudette! All examiners did agree that the correct hair was consistent with the known sample. The hairs that caused the type II errors in this experiment were common featureless hairs.

In the second experiment, Gaudette compared 100 known hair samples to one questioned hair. He repeated the experiment three times using different sets of hairs. Twice the one and only correct known sample was picked as being consistent with the questioned hair (i.e. no type I or type II errors were made). In the third trial, a common featureless hair was chosen as the questioned hair. This hair was found to be consistent with two of the known samples, the correct one and one other (i.e. one type II error was made).

Strauss in 1983 conducted a series of seven experiments in which 10 questioned hairs were compared to 10 known samples. In each of the seven experiments, the known and questioned hairs were selected by a neutral party from a hair pool so that different numbers of questioned hairs actually matched the known samples. Each time Strauss correctly matched all questioned hairs to their correct known samples (i.e. no type I or type II errors were made).

Bisbing and Wolner reported on a study in 1984 whereby each of seven questioned hairs were compared to several known samples. The results are shown in **Table 1**. Hairs in this study were from twins, the majority of whom were below the age of six. The majority of the subjects were blond. Most of the hairs were common featureless types and cut samples were used, thereby reducing the number of comparative features.

Wickenheiser and Hepworth repeated the Gaudette and Keeping study in 1990, with experimental modifications designed to overcome some of the criticisms of the original study. Wickenheiser and Hepworth collected representative hair samples of at least 100 hairs from each of 97 Caucasian individuals, including some closely related people from several generations. They then selected 5–13 hairs from each sample as representative of the range of characteristics present. The principal variation from Gaudette and Keepings' procedure was that they had an independent person randomly number the mutually dissimilar hairs and add 53 additional hairs randomly chosen from the original known samples of the 97 individuals. By including several duplicate hairs in the study, Wickenheiser and Hepworth ensured that if they encountered hairs they could not distinguish, they would not be biased by the knowledge that the hairs had to have originated from different sources.

With the assistance of a personal computer database to eliminate unnecessary microscopic comparisons of obviously dissimilar hairs (one of the authors

Table 1 Results of simulated forensic comparisons

Questioned specimen number	*Hair color*	*Number of known specimens*	*Number of matches*[a]		*Comments*
			Examiner 1	*Examiner 2*	
1	Brown	10	1	0	Hairs differ
2	Blond	10	2	2	Known contained duplicates
3	Blond	5	0	0	
4	Brown	5	0	0	Twin eliminated
5	Brown	7	0	0	
6	Blond	7	1	0	Indistinguishable in liquid media
7	Blond	8	1	1	

[a] All matches are incorrect (type II errors). No type I errors were made because none of the known pools contained the true source. Reproduced with permission from Bisbing RE and Wolner MF (1984) Microscopical discrimation of twins head hair. *Journal of Forensic Sciences* **29**: 780–786.

still made 749 one to one microscopic comparisons, and the other author required 2006), they were able to conduct 431 985 pairwise hair comparisons. One author found seven pairs of hairs to be indistinguishable, and the other author found six. In all cases these matches were between duplicate hairs; neither examiner found any hairs from different individuals which coincidentally matched. Of the 53 duplicate hairs, 38 were found to be unique in that they had no matching hair in the known sample selected.

This study led to several interesting conclusions.

1. If a one to one microscopic match is found between two hairs, the chances of it being a coincidental match are remote;
2. As reflected in the differences between the two examiners with respect to the number of direct microscopic comparisons required, the classification of hairs varies greatly between examiners;
3. The classification of hair is inconsistent due to variations over time. This then made the sorting procedure susceptible to error;
4. Five to thirteen macroscopically selected hairs are frequently inadequate to represent a known sample. This is the reason they did not find more matches between duplicate hairs. This led the authors to conclude that experimental work aimed at determining the optimum composition of a representative known hair sample is warranted.

The only reported study of the significance of non-Caucasian hairs was by Lamb and Tucker in 1994. In connection with the investigation of a series of sexual assaults, they compared known samples from 118 Afro-Caribbean suspects to questioned hairs from three crime scenes. Because the samples were collected over a two-year period, a full range of characteristics (such as length) could not be used and the level of discrimination had to be downwardly adjusted. Nevertheless, they were able to eliminate 62% of the suspects through low power microscopic examination with incident light. A further 25% of suspects were eliminated by transmitted light microscopy at higher powers, leaving only 9% of suspects which could not be eliminated.

Apart from Gaudette's pubic hair study, no study on the value of non-scalp human hair comparisons has been published.

A study of the significance of dog hair comparison was conducted in conjunction with a celebrated American murder case (*State of Georgia* v. *Wayne Williams*). Gaudette compared the hairs from the suspect's German shepherd dog to hairs from 12 other German shepherd dogs. The hairs from the suspect's dog were divided into ten types depending on color and whether they were guard or intermediate hairs. Eight of the twelve comparison dogs had no hairs matching any of these ten types. Three of the comparison dogs each had one type of hair that was macroscopically and microscopically indistinguishable from one type of hair on the suspect's dog. The remaining dog had two hair types indistinguishable from the suspect's dog. It should be noted that the 12 comparison dogs were not selected at random from the population of all dogs but were deliberately chosen so that their coats closely matched the suspect's dog. If they had been randomly chosen, an even smaller number of coincidental matches would have been found.

In a blind study involving comparison of 15 questioned hairs to known hair samples obtained from 25 pure bred German shepherd dogs, no type II errors were made and 6 of the 15 questioned hairs were correctly assigned to their known sample of origin. In

a later extension of this study, a comparison of 25 questioned hair samples of about 10 hairs each to known samples from 100 mixed breed and purebred dogs of various types resulted in all 25 being correctly assigned with no incorrect associations.

From these various studies the following can be concluded about the value of forensic hair comparison evidence.

1. With a few isolated exceptions, hairs are not unique to an individual. Accordingly, it is possible for type II errors due to coincidental matches to occur in forensic hair comparison.
2. Type II errors are a relatively rare event in forensic hair comparisons conducted carefully by qualified, well-trained examiners. Accordingly, hair comparison evidence is generally good corroborative evidence.
3. There are several factors which can increase or decrease the probability of type II errors in a given case. Accordingly, each case must be considered on its own merit.

Use of Frequency of Occurrence Data

Some forensic scientists have proposed setting up a computerized database of hair comparison characteristics which they would then use to state frequency of occurrence data in court. There are, however, many problems with such an approach. First, presentation of frequency data on its own can lead to a distorted picture of the value of evidence along with a false sense of exactness. Second, there is the difficulty of characterizing the hairs for a database. It requires examiners to adopt a check list approach rather than the more natural pattern recognition approach. Two hairs described as alike can be markedly different microscopically. Two examiners are likely to describe hairs in slightly different ways. The same examiner will even vary his or her description from day to day. And finally, setting up such a database would be extremely time consuming. Accordingly, results-oriented research (such as previously described studies) is much preferable to the database approach. This is not to suggest, however, that information from databases would not be valuable. On the contrary, it could be quite useful in helping examiners decide which characteristics, and combinations thereof, are unusual and which are common.

Report Writing and Court Testimony

On the basis of the results of an examination, the hair examiner must draw a conclusion which he or she then interprets in giving an expert opinion as to evidential value. Conclusions and expert opinions are given in report writing and court testimony. Exact wording of conclusions will depend on an examiner's preferences and a laboratory's policy. A symmetrical spectrum of conclusions such as the following is suggested. (A positive conclusion is defined here as one drawn from a finding of similarity between a known sample and a questioned hair. A negative conclusion is one arising from a finding of dissimilarity.)

Strong positive:	The questioned hairs originated from the same person as the known sample.
Normal positive:	The questioned hairs are consistent with having originated from the same person as the known sample.
Inconclusive:	No conclusion can be given as to whether the questioned and known hairs has a common origin.
Normal negative:	The questioned hairs are not consistent with having originated from the same person as the known sample.
Strong negative:	The questioned hairs could not have originated from the same person as the known sample.

The great majority of hair comparisons will result in normal positive or normal negative conclusions, with the other three being rarely encountered.

The normal positive and normal negative conclusions cover a wide range of evidential value. Accordingly, it is important that they be further interpreted in reports and court testimony. The examiner should first mention that hair comparison is not usually a positive means of personal identification. An estimate of the average value of forensic hair comparison evidence should then be given. This can be either based on personal experience or some of the previously described published studies. Factors weakening or strengthening the evidence in the particular case should then be mentioned. Some factors which can weaken hair evidence in a particular case are given in **Table 2**. Some factors which tend to strengthen normal positive hair comparison conclusions are given in

Table 2 Some factors which tend to weaken positive hair comparison conclusions

1. The presence of incomplete hairs.
2. Questioned hairs which are common featureless hairs.
3. Hairs of non-Caucasian racial origin.
4. A questioned hair found in conjunction with other unassociated hairs.
5. Known samples with large intrasample variation

Table 3 Some factors which tend to strengthen positive hair comparison conclusions

1. Two or more mutually dissimilar hairs found to be similar to a known sample.
2. Hairs with unusual characteristics.
3. Hairs found in unexpected places.
4. Two way transfer, for example, a victim's hair found on an accused's clothing and an accused's hair found on the victim's clothing.
5. Additional examinations

Table 4 Some factors which tend to weaken normal negative hair comparison conclusions

1. Deficiencies in the known sample.
   (a) not enough hairs,
   (b) not representive,
   (c) contains incomplete hairs,
   (d) large time difference between offence and procurement of known sample.
2. Incomplete questioned hairs.
3. Questioned hair has macroscopic characteristics close to those of the known sample.

Table 5 Some factors which tend to strengthen normal negative hair comparison conclusions

1. Known sample has more than the recommended number of hairs.
2. Known sample shows little intrasample variation.
3. Questioned hair has macroscopic and microscopic characteristics very dissimilar to those of the known sample.
4. Two or more questioned hairs found together in a clump are dissimilar to the known sample.

Table 3. Similarly, for negative conclusions, factors weakening them are given in **Table 4** and factors strengthening them are given in **Table 5.**

See also: **Evidence**: Statistical Interpretation of Evidence/Bayesian Analysis. **Expert Witness**: Qualifications and Testimony. **Hair**: Overview; Hair Transfer, Persistence and Recovery; Identification of Human and Animal Hair; Comparison: Microscopic; Comparison: Other; Deoxyribonucleic Acid Typing.

Further Reading

Barrett P and Ogle R (1982) Probabilities and human hair comparison. *Journal of Forensic Sciences* **27**:272–278.

Bisbing RE and Wolner MF (1984) Microscopic discrimination of twins head hair. *Journal of Forensic Sciences* **29**:780–786.

Camps FE (ed.) (1968) *Gradwhohls's Legal Medicine* 2nd edn, pp. 222–230. Bristol: John Wright.

Gaudette BD (1976) Probabilities and human pubic hair comparisons. *Journal of Forensic Sciences* **21**:514–517.

Gaudette BD (1978) Some further thoughts on probabilities and human hair comparisons. *Journal of Forensic Sciences* **23**:758–763.

Gaudette BD (1982) A supplementary discussion of probabilities and human hair comparison. *Journal of Forensic Sciences* **27**:279–289.

Gaudette BD (1985) Strong negative conclusions in hair comparisons – a rare event. *Canadian Society of Forensic Sciences Journal* **17**:32–37.

Gaudette BD (1986) Evaluation of associative physical evidence. *Journal of Forensic Science Society* **26**:163–167.

Gaudette BD (1985) A collaborative study in forensic hair comparison. *Proceedings of the Tenth International Association of Forensic Sciences, Oxford 1984.* Solihull, UK: AC Associates.

Gaudette BD and Keeping ES (1974) An attempt at determining probabilities in human scalp hair comparison. *Journal of Forensic Sciences* **19**:599–606.

Hoffman K (1991) Statistical evaluation of the evidential value of human hairs possibly coming from multiple sources. *Journal of Forensic Sciences* **36**:1053–1058.

Kirk P (1940) Human hair studies 1. General considerations of hair individualization and its forensic importance. *Journal of Criminal Law and Criminlogy* **31**:486–496.

Lamb P and Tucker L (1994) A study of the probative value of Afro-Caribbean hair comparisons. *Journal of the Forensic Science Society* **34**:177–179.

Martin E (1957) Further information on hair as a means of identification. *International Criminal Police Review* 303–305.

Polodak A and Blythe C (1985) A study of the feasibility of establishing a computer data bank of hair characterization using standard descriptive criteria. *Proceedings of the International Symposium on Forensic Hair Comparisions* FBI Academy, Quantico Virginia.

Strauss M (1983) Forensic characterization of human hair I. *The Microscope* **31**:15–29.

Suzanski TW (1988) Dog hair comparison: a preliminary study. *Canadian Society of Forensic Sciences Journal* **21**:19–28.

Suzanski TW (1989) Dog hair comparison: purebreds, mixed breeds, multiple questioned hairs. *Canadian Society of Forensic Sciences Journal* **22**:1323–1329.

Wickenheiser R and Hepworth D (1990) Further evaluation of probabilities in human scalp hair comparisons. *Journal of Forensic Sciences* **35**:1323–1329.

Deoxyribonucleic Acid Typing

M Yoshino, H Sato, National Research Institute of Police Science, Kashiwa, Chiba, Japan
S Seta, Department of Legal Medicine, St Marianna University School of Medicine, Kawasaki City, Japan

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0738

Introduction

Hair is commonly encountered as physical evidence in a wide variety of crimes. Human hair is understandably of special importance in criminal cases involving personal contact, such as homicide and sexual assault. In some cases, animal hair can also be important.

Since hair continually falls from the body of every person it is often present at the crime scene or on the clothing of the victim and suspect. In general, hair is not easily destroyed except for its decomposition by microbial or insect invasion, and even after soft tissue decomposition, hair remains useful for personal identification and comparison for a long period of time.

Any review of the forensic aspects of hair examination must start with the observation that it is not yet possible to individualize a human hair to any single head or body. This does not imply that hair has no value as physical evidence. At present the initial method used for forensic hair examination is microscopical comparison of questioned hairs collected from the crime scene to standard hairs of the victim and suspect. In general, 25 standard hairs from at least five locations of the scalp would be collected from all individuals involved in the case as representative samples of an individual, since most head hairs are variable and of different color shade. In some laboratories, additional examinations such as ABO blood typing, elemental analysis and electrophoretic analysis of hair proteins have been carried out for characterizing hair samples. When hairs are properly collected at the crime scene and their submission to the laboratory is accompanied by an adequate number of control hairs, and the comprehensive examinations described above are performed, hair can provide strong corroborative evidence for placing an individual at a crime site. In order to enhance the degree of certainty of identification, DNA typing has been applied to forensic hair comparison.

Nuclear DNA in Hair

DNA is housed within the nucleus of all types of cells in the human body, except mature red blood cells. Hair is an appendage of the skin that grows out of a histological organ known as the hair follicle and naturally contains DNA. Hair growth stage is an important factor for extracting DNA from hair.

Hair grows cyclically with alternating periods of growth and quiescence. The hair follicle has three distinct growth phases. During growth, the follicle is in the anagen phase (**Fig. 1**). In the anagen phase, mitotically active cells above and around the dermal papilla of the follicle grow upward to form the medulla, cortex, cuticle and inner root sheath of hair. The subsequent resting state is called the telogen phase (**Fig. 2**). In the telogen phase, the hair is anchored in the follicle by only a club. The transition period, when follicle is moving into an inactive quasiembryonal state, is called the catagen phase. On a healthy head hair, one would expect to find 80–90%

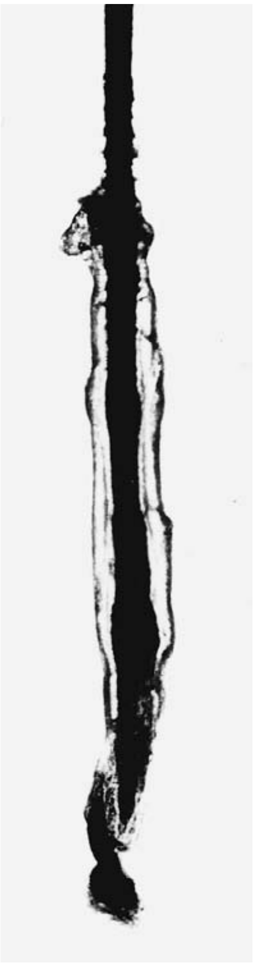


Figure 1 An anagen human hair root forcibly pulled from the scalp.

Figure 2 The telogen root of a human hair which fell from the scalp.

of the hair follicles are at the anagen phase of the growth cycle, 2% at the catagen phase, and 10–18% at the telogen phase. If hairs were forcibly pulled from the head of a criminal during a struggle and remained at the crime scene, hair samples as physical evidence would be in the anagen phase (plucked hair), containing many cells of the inner root sheath. From such plucked hairs, it is easy to extract the DNA by the same procedure as used for tissue samples because a greater part of DNA in hair is located in the root and surrounding sheath cells. A single plucked hair may contain as much as 100–500 ng of DNA. However, in most crime cases, human hair samples recovered from the scene of a crime are in the telogen phase (shed hair) which has naturally fallen out. In a hit-and-run case, only the hair shaft without the root region may be collected from a suspect car and submitted to the laboratory as physical evidence. In such a case, the DNA would have to be extracted from the nuclear remnant of the club and hair shaft (**Fig. 3**).

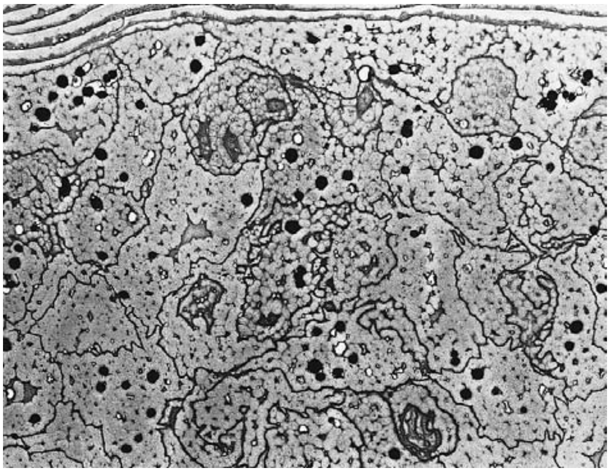


Figure 3 Transmission electron micrograph of human head hair. Each cortical cell contains nuclear remnants and microfibrils. Melanin granules are scattered throughout the cortex.

DNA extraction from hair

The generic procedure for DNA isolation from evidence samples is to digest the sample with a TNE buffer (10 mM Tris-HCl, 100 mM NaCl and 10 mM EDTA, pH 8.0) containing proteinase K followed by phenol/chloroform extraction. It is difficult to isolate DNA from shed hair by routine procedures because the shed hair contains very small and extremely degraded DNA. In order to effectively extract DNA from the shaft and root of this type of hair, keratinized cells of the hair must be completely broken down by the digestion buffer. Although the standard TNE buffer leads to partial dissolution of the hair shaft, a two-step digestion method using TNE buffer facilitates its dissolution. A digestion buffer containing calcium as activator of proteinase K also leads to total dissolution of the hair.

The DNA in the phenol/chloroform extract must be separated from the extraction solvents and concentrated prior to DNA amplification. For concentration and purification of DNA, several techniques such as the CTAB (cetyltrimethylammonium bromide) treatment, the GENECLENE (Bio 101) and the Microcon (Amicon) have been tested. In general, the effective concentration of DNA is performed by using spin ultrafiltration concentrators such as the Microcon equipment.

The procedure of DNA extraction from the shaft or root of shed hair is as follows.

A single shed hair is washed in a detergent (1% Tween-20), deionized sterilized water and absolute ethanol. The hair sample is cut at the level of 5 mm from the root and divided into the shaft and root. Five cm from the proximal end of the hair shaft is used for

DNA extraction. The hair shaft is cut into 5–10 mm pieces with a sterilized blade, and placed in a 1.5 ml microcentrifuge tube and then incubated in a digestion buffer (320 μl of TNE buffer, 40 μl of 10% SDS (sodium dodecyl sulfate), 40 μl of 0.4 M DTT (dithiothreitol) and 15 μl of 10 mg ml^{-1} proteinase K) at 55°C for 1 h. Then a second incubation is performed at 55°C for 1 h after addition of 20 μl of 10% SDS, 20 μl of 0.4 M DTT and 5 μl of 10 mg ml^{-1} proteinase K.

The DNA solution is then washed three times with TE-saturated (10 mM Tris-HCl and 1 mM EDTA) phenol (pH 8.0), once with phenol chloroform, once with chloroform, and once with ether. The DNA in the aqueous phase is then purified and concentrated using Microcon 100 ultrafiltration. Extraction of DNA from the hair root (3 mm in length) is performed by the same method.

In some reports, the amount of DNA in the hair shaft usually cannot be amplified to detectable levels with 30–35 cycles of PCR (polymerase chain reaction). However, according to one study, the DNA extracted from a 5 cm length of head hair shaft was about 20 ng and its base pair size was demonstrated to be lower than 500 bp (base pairs), mainly 100 bp by agarose gel electrophoresis.

Amplification and typing of products

PCR amplification PCR is a process by which short segments of DNA sequence can be selectively replicated a millionfold or more. The PCR process consists of three major steps: denaturing, annealing and extension. The denaturing step is the separation of the two strands of the DNA molecule. The double-stranded DNA is dissociated into single strands by incubation at 94–96°C. The two strands serve as templates for the replication of their complementary sequences. Primer binding is called annealing. The temperature is lowered (55–72°C) to allow the oligonucleotide replication primers to bind to their complementary sequences on the template DNA strands. In the next step (72–75°C), the primers are extended by the stepwise addition of nucleotides to the 3′ end of the primer strand using the target sequence as template. This step is mediated by DNA polymerase.

For hair shaft samples, 5 μl of DNA solution (corresponding to 2.5 cm of hair shaft length) is used as a template DNA. It has been reported that melanin contained in the hair is often extracted together with DNA and acts as an inhibitor of PCR amplification (**Fig. 4**). When melanin is present at concentrations higher than 1.6 ng μl^{-1} in the PCR solution, PCR amplification was significantly inhibited (Fig. 4). However, most melanin is preferentially removed from the extract by the phenol/chloroform extraction procedure.

DNA typing A number of DNA typing systems after the PCR amplification are available: AmpliFLP (amplified fragment length polymorphism); STRs (short tandem repeats); HLA DQ α and Polymarker (PM). In AmpliFLP analysis (D1S80, D17S30 etc.), the template DNA is amplified using primers which flank a core repeat of 8–16 bp. Fragment length polymorphism, like VNTR (variable number tandem repeat) polymorphism, is based on the number of core repeats found in a particular allele. AmpFLP alleles range in size from 100 to 1300 bp, whereas VNTR alleles range from 500 to 23 000 bp. STRs are similar to the AmpFLP system, except that the core repeat is shorter. STR alleles range from 100 to 400 bp and

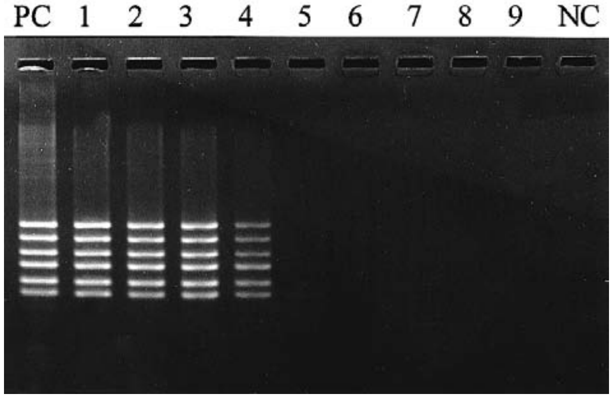


Figure 4 Influence of melanin on PCR of PM kit. PC: Positive control (melanin free); 1: 0.1 ng μl^{-1} melanin in PCR solution; 2: 0.2 ng μl^{-1}; 3: 0.4 ng μl^{-1}; 4: 0.8 ng μl^{-1}; 5: 1.6 ng μl^{-1}; 6: 3.2 ng μl^{-1}, 7: 6.4 ng μl^{-1}; 8: 12.8 ng μl^{-1}; 9: 25.6 ng μl^{-1}; NC: negative control.

their core repeats are 4–6 bp. In HLA DQ α and PM, the amplified DNA products are typed by dot-blot analysis. The probes detect six common DQ α alleles which, in combination, determine 21 possible genotypes. PM consists of five loci, low density lipoprotein receptor (LDLR, 214 bp), glycophorin A (GYPA, 190 bp), hemoglobin gammaglobin (HBGG, 172 bp), D7S8 (151 bp) and group specific component (GC, 138 bp). PM typing has been developed in which several markers at different loci are analyzed at the same time by the same technique used in HLA DQ α analysis.

Because the size of the DNA extracted from shed hair may be 100–500 bp, the loci with a small amplification size have to be used as a target for DNA typing of this type of hair. For this reason, it seems that STRs and PM typing are appropriate methods for routine hair comparison work.

There are approximately 4.0 × 10^8 STR loci dispersed throughout the human genome; SRT loci, TH01, CSF1PO, VWA, FGA, TPOX etc. are used for biological evidence. The amplification product is subjected to a polyacrylamide gel electrophoresis and then visualized by silver staining. A number of STR loci may be amplified and separated simultaneously, a technique known as multiplexing. Multiplex PCR kits such as Blue kit (FGA, vWA, D3S1358), Green kit (CSF1PO, TPOX, TH01) and Profiler (FGA, vWA, D3S1358, CSF1PO, TPOX, TH01, D7S820, D13S317, D5S818, PE/Applied Biosystems) are currently commercially available. The amplification products are electrophoresed and analyzed on a fragment analyzer. The combination use of STR loci enormously enhances the discrimination power of individuals, whereas in the severe degraded samples some loci of the multiplex system cannot be detected. For typing of DNA extracted from the hair, a single locus amplification may be recommended. TH01 and PM typing for the shed hairs are shown in **Figs 5** and **6**, respectively. TH01 locus was amplified using Quick-Type® HUMTH01 (Lifecords). PM typing was performed using AmpliType PM PCR Amplification and Typing kit (Perkin-Elmer). In PM typing, it is generally recommended that a DNA probe strip with no visible 'S' dot should not be typed for any locus ('S' dot is a minimum dot intensity control and corresponds to the 'C' dot on the HLA DQ α DNA probe strip). However, by comparing all bands of amplified DNA sample with relevant bands of amplified control DNA (template: 2 ng, provided in the kit) by agarose gel electrophoresis, all five or a few loci could be correctly typed even when the 'S' dot was not visible on the DNA probe strip. The success rate for DNA typing of the shaft and root of shed hair is estimated at about 25% and 50%, respectively. This suggests that the hair root is more suitable for STRs and PM typing when shed hair is the evidential sample. However, even in the root of shed hair, about 50% of samples cannot be analyzed by these DNA typing systems because of its limited and degraded nuclear DNA.

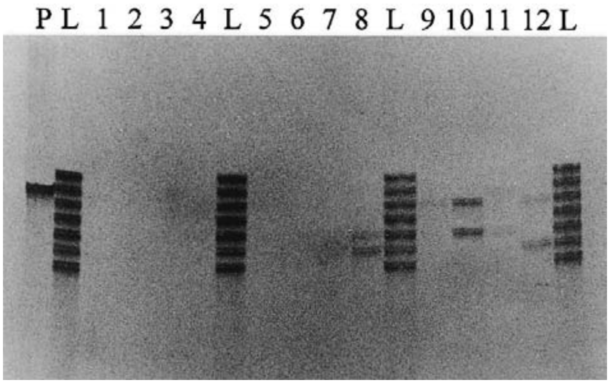


Figure 5 TH01 typing from head hair shaft of the telogen phase on 4% denatured polyacrylamide gel electrophoresis. P: positive control (K562 DNA); L: allelic ladder; 1–12: subjects 1–12; 8: 6–7; 10: 7–9; 12: 6–9.

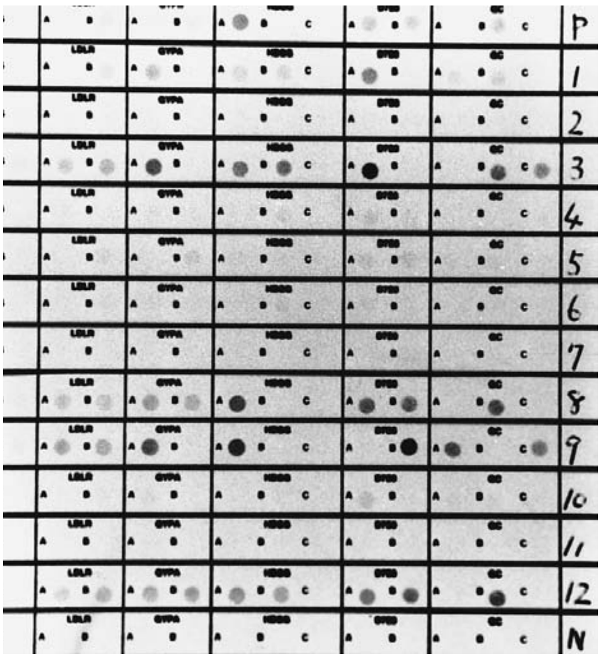


Figure 6 PM typing from head hair root of the telogen phase according to the protocol in the PM kit. P: positive control (K562 DNA); 1–12: subjects 1–12; N: negative control, 1: LDLR; BB, GYPA; AA, HBGG; AB, D7S8; AA, GC; AB, 3: AB, AA, AB, AA, BC, 8: AB, AB, AA, AB, BB, 9: AB, AA, AA, BB, AC, 12: AB, AB, AB, AB, BB.

Mitochondrial DNA

In addition to nuclear DNA, eukaryotic cells contain a unique mitochondrial form of DNA. The inheritance of mitochondrial genes is exclusively maternal, and clearly non-Mendelian. Mitochondria in human cells contain between 2 and 10 copies of a double-stranded, circular DNA molecule that is about 16.5 kb in length. Owing to the high copy number of mitochondrial DNA molecules per mitochondrion and the high number of mitochondria per cell, there is 1000–10 000-fold molar excess of mitochondrial DNA (mtDNA) compared to nuclear DNA.

Hypervariable regions

The mitochondrial genome contains a noncoding hypervariable control region. In particular, there are two segments, hypervariable region 1 (HV1:16024-16365) and hypervariable region 2 (HV2:73-340) within this 1100 bp control region (**Fig. 7**), also called the D-loop, that tend to mutate with an extremely high frequency, at least 5–10 times that of nuclear DNA. This high copy number and the high mutation rate of mtDNA are the main reasons for its use in individual identification. The complete 16 569 nucleotide sequence of human mtDNA has been established for a reference individual. For present, all comparisons are made to this reference called the Anderson sequence.

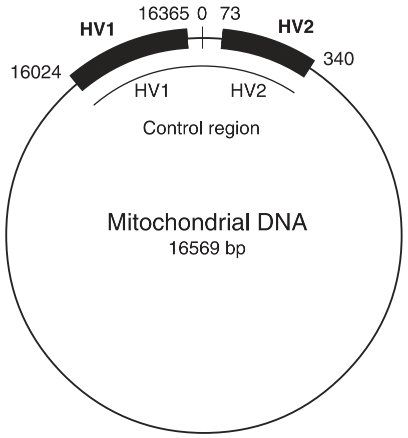


Figure 7 Locations of the two HV regions on the mtDNA circle.

Amplification and sequencing of mtDNA

PCR amplification In general, HV1 and HV2 in control region are subjected to mtDNA analysis. Several protocols for the PCR amplification of mtDNA control region and sequence analysis of the PCR product have been established.

Two sets of PCR primers are used to amplify overlapping segments from both HV1 and HV2. Examples of PCR primers used for the hair sample are shown in **Table 1**. The primer pairs used to amplify HV1 and HV2 are: HV1-A; A1/B2, HV1-B; A2/B1, HV2-A; C1/D2, HV2-B; C2/D1.

Amplification is performed in a reaction mixture of 25 μl using a thermal cycler. Each reaction mixture comprises 50 mM KCl, 10 mM Tris-HCl (pH 8.3), 15 mM $MgCl_2$, 10.0 pmol of each primer (0.2 μM), 200 μM each dNTP, 2.5 U Taq polymerase (AmpliTaq Gold, Perkin-Elmer) and 5 μl of the hair extract DNA (corresponding to 5 mm of hair shaft length). Thermal cycling conditions are 9 min at 95°C, followed by 36 cycles of 95°C for 10 s, 54°C for 30 s and 72°C for 30 s. A portion of the PCR products are taken to verify the amplification of the expected fragment by agarose gel electrophoresis and ethidium bromide staining prior to DNA sequencing (**Fig. 8**).

DNA sequence DNA sequencing is performed on mtDNA PCR products using fluorescence-based chemistry. In general, two types of fluorescent dye methods, namely a dye terminator cycle sequencing and dye primer cycle sequencing are used. The PCR products must be purified prior to DNA sequencing to remove unincorporated PCR primers and excessive dNTPs. The purified PCR products are quantified using a fluorometer, and diluted with deionized

Table 1 Examples of primer set for hair samples

Primer set	*Forward primer light strand*	*Reverse primer heavy strand*
HV1-A	L15997 5′-CACCATTAGCACCCAAAGCT-3′	H16236 5′-CTTTGGAGTTGCAGTTGATG-3′
HV1-B	L16159 5′-TACTTGACCACCTGTAGTAC-3′	H16401 5′-TGATTTCACGGAGGATGGTG-3′
HV2-A	L048 5′-CTCACGGGAGCTCTCCATGC-3′	H285 5′-GGGGTTTGGTGGAAATTTTTTTG-3′
HV2-B	L172 5′-ATTATTTATCGCACCTACGT-3′	H408 5′-CTGTTAAAAGTGCATACCGCCA-3′

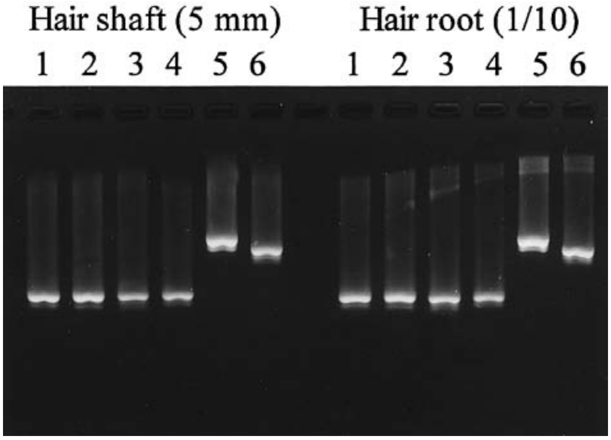


Figure 8 Agarose gel electrophoresis of PCR of mtDNA extracted from head telogen hair. Template DNA corresponds to the amount of DNA extracted from a 5 mm length of the hair shaft and one-tenth of the hair root. Primer pair: 1; A1/B2, 2; A2/B1, 3; C1/D2, 4; C2/D1, 5; A1/B1, 6; C1/D1.

water to a suitable concentration. For example, sequencing fragments are generated using the dye terminator cycle sequencing kit (Applied Biosystems). Following gel filtration, the purified extension products are dried in a vacuum centrifuge. The purified products are electrophoresed on a 5–6% polyacrylamide sequencing gel on an AB 373A DNA sequencer or 377 DNA sequencer and the AB computer software collects and analyzes the raw data.

Sequence data interpretation When the sequencing is successfully achieved from both forward and reverse direction (**Fig. 9**), the sequences are aligned and compared with the reference sequence using the Sequence Navigator computer program (Applied Biosystems). Each nucleotide substitution, deletion or insertion is registered. Polymorphism of mtDNA in an individual is determined based on the difference of nucleotides at some position between the sequence obtained from the individual and the reference sequence (**Fig. 10**). However, the sequences cannot be determined in some individuals who have an anomaly called heteroplasmy within the mtDNA control region. There are two kinds of heteroplasmy in human mtDNA; sequence heteroplasmy and length heteroplasmy. Sequence heteroplasmy is a condition in which more than a single base is observed as mixed peaks at a particular position in electropherogram. This indicates the presence of two variant populations of mtDNA in the same individual. In length heteroplasmy, because of the existence of more than two variants in length in a homopolymeric tract of cytosines between nucleotides 16184 and 16193, the sequences of 3′-side lying ahead of this tract are hardly determined. Because the presence of heteroplasmy might potentially lead to a false exclusion in a forensic case, careful analysis and comparison of known and questioned samples is necessary.

Currently the mtDNA is being used forensically, and its success rate for analysis of questioned samples is over 80%. At the present, mtDNA analysis is the most effective method for hair sample to which nuclear DNA typing cannot be applied.

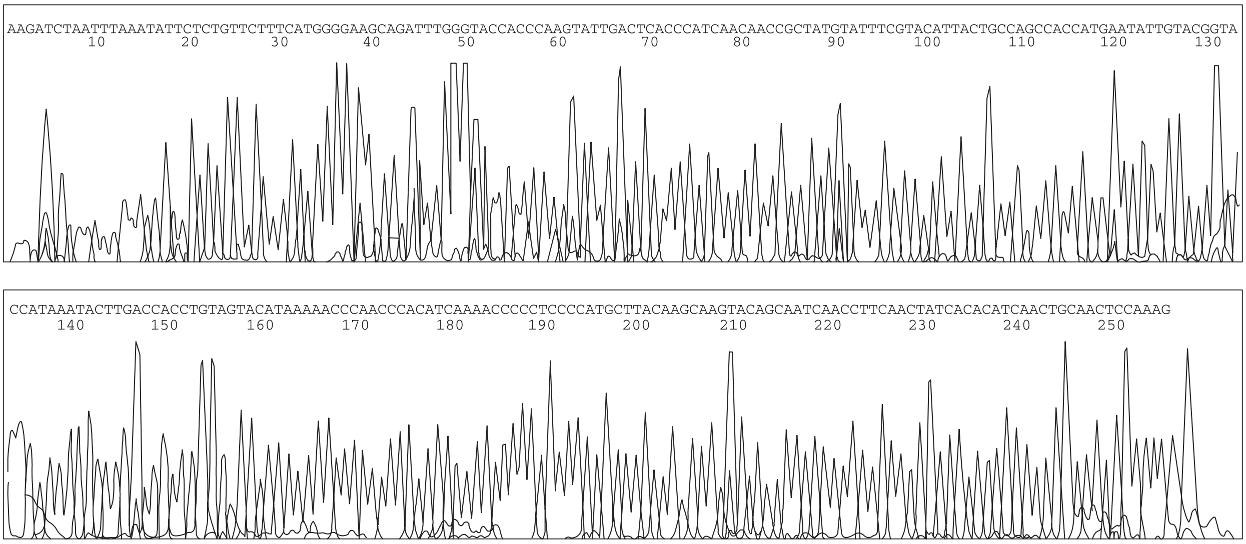


Figure 9 Electropherogram for the forward HV1-A (Primer A1) analyzed by ABI PRISM™ Sequencing Software Ver.2.1.2.

	16250	16260	16270	
REF	ACTGCAACTC	CAAAGCCACC	CCTCACCCAC	TAGGATAC
A	ACTGCAACTC	CAAAGCCACC	**T**CTCA**G**CCAC	TAGGATAC
B	ACTGCAAC**C**C	CAAAGCCACC	CCTCACCCAC	TAGGATAC
C	ACTGCAACTC	CAAAGC**A**ACC	**T**CTCACCCAC	TAGGATAC
D	AC**C**GCAACTC	CAAAGCCACC	CCTCACCCAC	TAGGATA**T**
E	ACTGCAACTC	CAAAGCCACC	CCTCA**T**CCAC	TAGGATA**T**

Figure 10 A comparison between hair DNA obtained from five individuals and reference sequence. Pyrimidine substitutions (C-T or T-C transition) are found at 16243 (D), 16249 (B), 16261 (A and C), 16266 (E) and 16278 (D and E). C-A and C-G transversions are observed at 16257 (C) and 16266 (A), respectively.

Factors Causing DNA Damage

Environmental insults

Hair samples collected from the scene of a crime may have been exposed to environmental conditions such as lighting, high temperature and humidity, dust, soil, bacteria and so on. These various environmental exposures have an effect on DNA analysis.

A major consequence of ultraviolet (UV) damage caused by lighting is the formation of covalent links between adjacent pyrimidines to make pyrimidine dimers. PCR product yields are reduced as the dose of UV exposure is increased because *Taq* polymerase stops at UV damage sites. If hair samples are exposed to lighting for a long period and received very high UV doses, PCR is completely inhibited. Microbial contamination occurs when hair samples are located in the dust or soil under the warm and moist conditions. In this case, degradation of DNA is caused by microbial activity which causes strand scission and results in the preferential loss of the larger allele products.

Cosmetic treatments

Hair color is altered by bleaching and dyeing, and form is changed by styling such as permanent waving. Any of these procedures may produce irreversible damage to the hair shaft.

Hair bleaching using hydrogen peroxide at alkaline pH selectively decolorizes the natural hair pigment. Permanent hair dye is a system of hair coloring in which the cortex of the hair is colored by reaction of the colorless dye precursors with hydrogen peroxide. In both hair bleaching and coloring, oxidation processes are used to break down melanin granules. Such oxidizing agents cause degradation of DNA in the hair shaft, such as base modification and strand scission.

In permanent waving, three stages are involved; physical and chemical softening of the hair, reshaping and hardening of fibres to retain the reshaped position. A permanent curl in the hair is accomplished by breaking the disulfide linkage of keratin with a reducing agent and reforming the pliable hair with a curling rod. The hair is then oxidized with agents such as hydrogen peroxide, sodium perborate or sodium percarbonate, so that the disulfide bonds are reconstructed and the curl fixed. It is generally accepted that organic solvents and reducing agents have no effect on the quality of DNA. However, the DNA in the hair shaft is damaged by oxidation in the last stage of permanent waving.

See also: **Deoxyribonucleic Acid**: Significance. **Hair**: Overview.

Further Reading

Bisbing RE (1982) The forensic identification and association of human hair. In: Saferstein R (ed.) *Forensic Science Handbook*, pp. 184–221. Englewood Cliffs, NJ: Prentice Hall.

De Forest PR, Gaensslen RE and Lee HC (1983) *Forensic Science, An Introduction to Criminalistics*, pp. 192–229. New York: McGraw-Hill.

Duncan GT and Tracey ML (1997) Serology and DNA typing. In: Eckert WG (ed.) *Introduction to Forensic Sciences*, pp. 233–293. New York: CRC Press.

Erlich HA (1989) *PCR Technology, Principles and Applications for Forensic Amplification*. New York: Stockton Press.

Farley MA and Harrington JJ (eds) (1991) *Forensic DNA Technololgy Science*. Michigan: Lewis Publishers.

Higuchi R, von Beroldingen CH, Sensabaugh GF and Erlich HA (1988) DNA typing from single hairs. *Nature* **332**:543–546.

Inman K and Rudin N (1997) *An Introduction to Forensic DNA Analysis*. New York: CRC Press.

Kobilinsky L (1993) Deoxyribonucleic acid structure and function – a review. In: Saferstein R (ed.) *Forensic Science Handbook*, vol. III, pp. 287–357. Englewood Cliffs, NJ: Regents/Prentice Hall.

Lee HC and Gaensslen RE (eds) (1990) *DNA and Other Polymorphisms in Forensic Science*. Chicago: Year Book Medical Publishers.

Matsuda H, Sekiguchi K, Kasai K, Yoshino M and Seta S (1997) Evaluation of various methods of DNA analysis for hair samples. *Japanese Journal of Science and Technology for Identification* **2**:79–84.

National Research Council (1992) *DNA Technology in Forensic Science*, Washington, DC: National Academy Press.

Saferstein R (1990) *Criminalistics, An Introduction to Forensic Science*, pp. 178–215, 343–362. Englewood Cliffs, NJ: Prentice Hall.

Sensabaugh GF and Blake ET (1993) DNA analysis in biological evidence: applications of the polymerase chain reaction. In: Saferstein R (ed.) *Forensic Science Handbook*, vol. III, pp. 416–452. Englewood Cliffs, NJ: Regents/Prentice Hall.

Seta S and Sato H (1988) Forensic hair investigation. In: Maehly A and Williams RL (ed.) *Forensic Science Progress*, vol. 2, pp. 47–166. Berlin: Springer.

Waye JS and Fourney RM (1993) Forensic DNA typing of highly polymorphic VNTR loci. In: Saferstein R (ed.) *Forensic Science Handbook*, vol. III, pp. 358–415. Englewood Cliffs, NJ: Regents/Prentice Hall.

Wilson MR, Stonneking M, Holland MM, DiZinno JA and Budowle B (1993) Guidelines for the use of mitochondrial DNA sequencing in forensic science. *Crime Laboratory Digest* **20**:68–77.

Zviak C (1986) *The Science of Hair Care*. New York: Marcel Dekker.

Hair Transfer, Persistence and Recovery

B D Gaudette, Royal Canadian Mounted Police, Ottawa, Canada

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0556

How do hairs come to occur as physical evidence? The answer lies in the concepts of hair transfer, persistence and recovery. In order to properly interpret the significance of forensic hair examinations, it is essential to have a thorough understanding of these concepts combined with knowledge of hair identification and comparison methods.

Hair Transfer

The fundamental principle of hair transfer is the Locard Exchange Principle, first formulated by Edmond Locard who stated: 'The microscopic debris that cover our bodies are the mute witnesses, sure and faithful, of all our movements and all our encounters'. This principle means that whenever two objects are in contact, there will always be a transfer of material from one to the other, even though in some instances the amount of material transferred may be too small to detect, or may be rapidly lost subsequent to transfer.

The original source of all transferred hairs is the human or animal body area from which they grew. As part of the normal hair cycle, a person loses 75–100 scalp hairs a day. Of course, this loss is not evenly distributed throughout the day. Most occurs during hair grooming – combing, brushing and washing. A struggle including hair pulling (as often occurs during crimes of violence) can greatly accelerate hair loss.

As is the case with transfer of other types of trace evidence, hair transfer is a very complicated process with many mechanisms and involves many variables. Two main types of hair transfer can be identified: direct transfer in which hairs direct from the original source are transferred; and indirect transfers which involve one or more intermediaries such as clothing, bedding, and household or automotive upholstery. Direct transfer is always primary transfer (defined as that transfer which occurs whenever a person transfers his or her own hair to an object, place or other person). Indirect transfer can be either primary or secondary. Secondary transfer arises when a person transfers hair that is not his or her own to an object, place or other person. Unlike primary transfer, secondary transfer does not imply direct association and can lead to falsely incriminating evidence.

Gaudette and Tessarolo conducted several experiments in a preliminary attempt to obtain information on hair transfer mechanisms and the role of some of the variables involved, as well as to obtain a preliminary indication of the extent and importance of secondary hair transfer. Indirect hair transfer mechanisms were found to be much more common than direct ones. They also found that secondary transfer of human scalp hair can and does exist in casework situations. The extent of secondary transfer was found to be extremely variable, being dependent on such factors as texture and fiber type of clothing worn, grooming habits of persons involved, and whether or not horizontal objects (such as upholstered chairs or car seats) used by several people, are involved. Their results showed that when secondary transfer does occur, it is most often first order secondary transfer, with only one source between the place where the hair is found and the original donor. Secondary transfer with two or more intermediates can occur, but was not found to be common.

The following conditions have been found to increase the likelihood of secondary hair transfer: involvement of unclean individuals with poor grooming habits, involvement of individuals with many prior interpersonal contacts, when one or more party is wearing rough textured or wool clothing, when horizontal surfaces are involved, and when there is contact with objects used by several different people. On the other hand, when those involved are clean,

well-groomed people, or are wearing smooth textured or tight fitting clothing which is neat and clean, the likelihood of secondary transfer is greatly reduced. Assessment of the chances of secondary transfer having occurred in a given case should also be based on the total number of hairs transferred and whether or not there has been a two-way transfer. In cases where a large number of hairs are transferred, it is unlikely that they would all have been secondarily transferred and it is unlikely that both transfers in a two-way transfer would be secondary.

Quill removed hairs from his clothing at the beginning and end of each work day for a 30-day period. He then compared the hairs removed to known samples from himself, his family and his co-workers. He found that the few hairs recovered from his clothing during the day were contributed by himself and those in his immediate environment.

Peabody, Thomas and Stockdale investigated the shedding of hairs into several types of headgear. They concluded that the number of hairs shed into headgear varies with the type of headgear and the individual, and that hairs from headgear are most similar to hairs from known samples obtained by combing.

In a study of pubic hair transfer during sexual intercourse, at least one transferred hair was found in 17.3% of combings taken from participants immediately after intercourse. About a quarter of these transfers involved multiple hairs (up to four). Transfers from females to males were found to be about twice as prevalent as transfers from males to females.

Hair Persistence

Since it is rare for suspects to be apprehended immediately after hair transfer in a crime, the concept of hair persistence on clothing becomes important. Gaudette and Tessarolo found the persistence of hairs on clothing to be quite similar to that previously found for fibers. A typical decay curve is shown in **Fig. 1**. Movement and loss of hairs on clothing has been found to be influenced by many variables and it was concluded that 'it would be quite unjustified to attempt to predict accurately the number of hairs likely to be present in a specific case or to attach great significance to where these hairs are found'. Although most human hairs are removed from items during laundering, some can still be found on items after laundering and hair transfers can occur during laundering.

Through a study of casework received over a six-year period, Mann obtained transfer and persistence results for scalp hairs that were consistent with those of the above mentioned studies. However, it was found that forensically significant pubic hair transfer occurred quite infrequently in the cases examined.

In addition to its ubiquitous nature, hair has two other properties that make it a frequently occurring type of physical evidence. First, it is remarkably stable to most environmental conditions and will not easily break down like other biological evidence. Second, since hair is fairly unnoticeable to the untrained eye, a criminal is not likely to make a special effort to destroy it.

Recovery of Hair Evidence

Recovery of hair evidence can occur in two locations: at the crime scene, and in the laboratory. Since the laboratory has the better facilities, as much hair recovery as possible should be conducted there. Accordingly, if an item suspected of having adhering hairs is at all portable, the best method of hair recovery at the scene is to seize and carefully package the entire item. Each exhibit item should be packaged immediately in its own separate container. Clean bags with no exposed inside seams are the best containers for clothing and other large items. Clear plastic vials or metal ointment tins are recommended for small objects. To prevent accidental contamination, exhibit items in an unpackaged state should not be handled by more than one investigator. By placing clear cellulose tape in contact with the area in question, hairs can be recovered from items which cannot practicably be brought to the laboratory.

In the laboratory, the simplest method of recovering hairs is to visually search items under oblique lighting at different angles, using tweezers to pick off any readily visible hairs. The examination table should be of a size sufficient to accommodate large items, have a smooth nonglossy white finish (to assist in finding hairs that fall to the table), be carefully

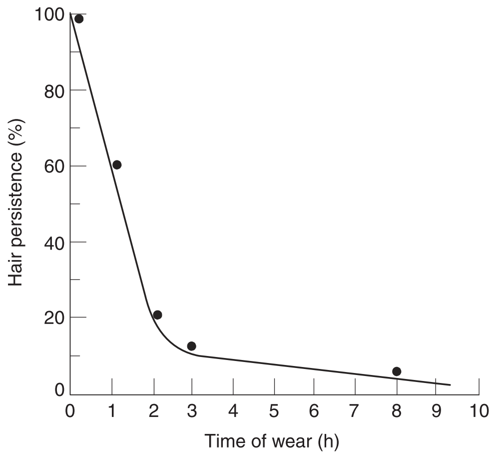


Figure 1 Hair persistence as a function of time of wear.

cleaned prior to each use, be located near a good source of natural lighting, and be equipped with a movable light source capable of providing illumination to the entire area.

Particularly with dark-colored and rough-textured items, a second method of hair recovery in addition to visual searching should usually be employed. The recommended method is taping. Although originally designed as a method to recover textile fibers, experience has shown that taping can recover many hairs which are not readily visible.

In the taping procedure, a strip of transparent cellulose tape is placed, adhesive side down, on the item and the back surface of the tape is rubbed with a thumb or forefinger. The process is systematically repeated until the entire item or area of interest has been taped. The tape strips are then placed between two layers of clear plastic (such as a clear acetate document protector).

Vacuuming is a method of hair recovery that is not recommended under most circumstances since it has some serious disadvantages. Most notable among these are the potential for contamination from hairs and debris lodged in the vacuum cleaner, and the nondiscriminating nature of the collection. Hairs that were deposited long before the crime are collected simultaneously with hairs associated with the crime, with no way to readily distinguish between them.

Since hairs are readily transferable, it is important to follow a rigid program of contamination prevention. A program similar to that followed for textile fibers or polymerase chain reaction DNA analysis is recommended.

See also: **Fibres**: Transfer and Persistence. **Hair**: Deoxyribonucleic Acid Typing.

Further Reading

Exline DL, Smith FP and Drexler SG (1988) Frequency of pubic hair transfer during sexual intercourse. *Journal of Forensic Sciences* **43**:505–508.

Gaudette BD and Tessarolo T (1987) Secondary transfer of human scalp hair. *Journal of Forensic Sciences* **32**:1241–1253.

Locard E (1930) The analysis of dust traces. *American Journal of Police Science* **1**:276–291.

Mann M (1990) Hair transfers in sexual assault: a six-year case study. *Journal of Forensic Sciences* **35**:951–955.

Moreau DM (1987) Major physical evidence in sexual assault investigations. In Hazlewood R and Burgess A (ed) *Practical Aspects of Rape Investigation*. New York: Elsevier.

Peabody AJ, Thomas K and Stockdale R (1985) On the nature of human hairs shed onto various types of headgear. *Proceedings of the Tenth International Association of Forensic Sciences, Oxford 1984*. Solihull, U.K: AC Associates.

Quill JL (1987) The transfer theory of hairs applied to the normal work day. *Proceedings of the International Symposium on Forensic Hair Comparison*. Washington, DC: Federal Bureau of Investigation.

Robertson J, Harding H and Somerset H (1987) The persistence of hairs on clothing. *International Association of Forensic Sciences Eleventh Triennial Meeting, Vancouver*.

Simons A (1986) Hair evidence on laundered items. *Crime Laboratory Digest* **13**:78–81.

Identification of Human and Animal Hair

B D Gaudette, Royal Canadian Mounted Police, Ottawa, Canada

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0559

Introduction

There are two main types of forensic hair examinations: hair identification and hair comparison. Hair identification involves classifying a hair as being a member of a specific group, e.g. human pubic hair. In hair comparison, on the other hand, a determination is made as to whether or not hairs of the same group are consistent with having had a common origin. Hair identification should be considered first since it is a necessary preliminary step in hair comparison.

Elements of Hair Identification

Six elements comprise the process of hair identification:

- the questioned hair;
- a standard reference collection;
- literature references including published photomicrographs and keys;
- the equipment used;
- the list of hair characteristics;
- the examiner.

The hair to be identified is termed the questioned hair. Unless there is some logical justification for grouping them together, all questioned hairs must be considered individually.

A good authenticated standard reference collection is a most important element in forensic hair

identification. Such collections, which should contain a large number of hair samples of different types and origins, can be built up through contacts with zoos and exchanges with other forensic scientists, or can be purchased commercially.

The next element in forensic hair identification is literature references. These include published photomicrographs and keys, which can supplement but not replace standard reference collections. Some of the citations in the bibliography provide published photomicrographs and keys.

The most important piece of equipment for hair identification is the comparison microscope. This consists of two compound microscopes joined by a common bridge. Using a mounting medium with refractive index of about 1.52, the questioned hair is mounted on a glass microscope slide and placed on the stage of one microscope. A microscope slide of a standard sample is placed on the stage of the other microscope. Both samples can then be viewed simultaneously in a split field. Comparison bridges also have an adjustment enabling either microscope to be used separately, if so desired. In forensic hair identification, magnifications of from 40× to 400× are generally used. Forensic hair examiners should be aware of the basic principles of microscopy.

Another useful piece of equipment for hair identification is a scanning electron microscope (SEM). This can be used to view hair scale patterns, cross-sectional shapes and medullary structures. Lack of access to an expensive piece of equipment like an SEM need not be a deterrent to those involved in hair identification. As will be seen in the following discussion, by means of some special techniques, the same information can be obtained using a light microscope and some inexpensive equipment.

Cast impressions of hair scale patterns can be produced by many different methods. The following method produces good quality scale impressions fairly quickly.

- Prepare a mixture of three parts colorless nail polish and one part isoamyl acetate.
- Place two or three drops of this mixture on a clean glass microscope slide and spread evenly, using the end of a second slide.
- Place a clean hair in the wet mixture.
- Dry for 10 min at room temperature, then remove the hair.
- The scale impression is now complete and ready for microscopical examination.
- The hair may then be washed in iso-amyl acetate and used for further examination.

A simpler method has been proposed whereby scale impressions can be seen on hairs fixed to glass microscope slides with clear cellulose tape. This method has the advantage of also revealing medullary structure.

Many methods have also been proposed for producing cross-sections of hairs. The recommended method uses a Schwartz fiber microtome. The hairs are first folded over several times so that cross-sections can be obtained from as many different areas along the shaft as possible. The bundle of hair is then packed with absorbent cotton and fitted tightly into the aperture on the top plate of the microtome. After trimming the bundle and assembling the microtome, the drum is turned so as to raise the bundle by about 50 μm. A small dab of collodion-like solution called Micolac is then applied to the top of the hair bundle and allowed to dry for about 5 min. The section is cut in one even motion with a single edged razor blade. With a little practice, this method can fairly quickly produce good quality cross-sections.

By examining longitudinally split hairs by SEM, medullary microstructure has been shown to greatly assist in animal hair identification. However, the same information can be obtained with a light microscope by hair splitting, followed by clearing the medulla with glycerine–alcohol and examination with an oil-immersion objective.

The fifth element of hair identification, the list of characteristics, arises from the basic structure of the hair. These characteristics are listed in the Appendix.

The final and most important element in hair identification is the examiner, whose role is to synthesize all the information gained from the various characteristics and develop a pattern. This requires a considerable amount of training and experience. There are no hard and fast rules. In the following discussion many qualifying phrases such as ‘generally’ or ‘usually are’ are used. The examiner must always be aware of exceptions. No one characteristic by itself is the most important in establishing identification. Rather, the general pattern of characteristics is important. In hair identification, an examiner should strive to maximize information content while maintaining an acceptable level of possible errors, thereby arriving at conclusions that are neither understated nor overstated.

Human or Animal?

With hair identification, it can first be determined whether a hair is human (**Fig. 1**) or animal (**Fig. 2**): The following are some of the distinguishing features.

Length: Human scalp hairs are longer than most animal hairs.

Diameter: Human hairs are usually in the range of 0.05–0.15 mm; animal hairs can be narrower or coarser.

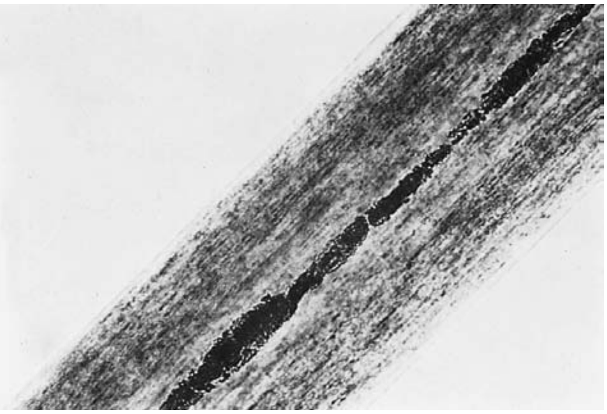


Figure 1 A human scalp hair (400×).

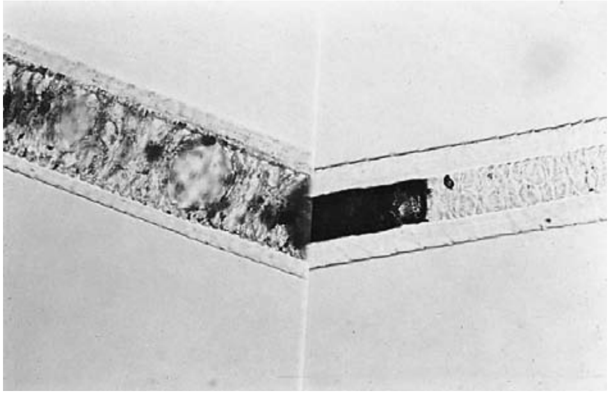


Figure 2 Left: Felidae (cat) hair; right: Canidae (dog) hair.

Color: Animal hair can have a banded appearance; untreated human hair never does.
Treatment: Except when used in textile materials, animal hair rarely exhibits dying, bleaching or other cosmetic treatment.
Medulla: Animal hairs can have a complex regular/geometric cellular medulla; human hairs have only an amorphous medullary structure.
Medullary index: Defined as the ratio of the diameter of the medulla to the diameter of the hair, the medullary index in human hair is almost always less than 1/3; in animal hair it is usually greater than 1/3.
Pigment distribution: Animal hairs can have pigment distributed about the medulla whereas this is extremely rare in humans.
Shaft: Animal hairs can have different shapes than human, e.g. the spatulate shape found in rodent hairs.
Root: Animal hairs have different root shapes, typically brush-like, whereas humans' are bulb or ribbon-shaped.
Tip: Human hair (particularly scalp) is usually cut or frayed at the tip; animal hair is generally naturally tapered.
Scales: Human hair only exhibits irregular annular scale patterns; animal hair has a variety of types and can have more than one type in the same hair.
Cross-section: Animal hairs can have some unusual shapes, e.g. dog bone (rabbit) and cigar shaped (seal).

Animal Hair Identification

The claim is often made in the literature that macroscopic and microscopic hair examination can usually result in species identification. It should be noted, however, that most of these books were written by nonforensic scientists. The consequences of error on the part of a forensic scientist are much greater than, for example, with someone doing wildlife habitat studies. Forensic scientists are also acting under the constraint that, unless found in groups, all questioned hairs must be considered singly. When this is added to the large variation in characteristics of hairs taken from different locations on the same animal and the wide animal to animal variation that can occur within a species (particularly in species spread over large geographical areas), it can be seen that forensic scientists should only macroscopically and microscopically identify species of animal hairs when good complete dorsal guard hairs are present and the examiner feels confident of the identification. In all other instances, only the animal family of origin should be identified.

The following are some major identifying features of animal hairs commonly occurring in forensic casework.

Canidae: Hairs usually have a complex cellular medulla with doughnut-shaped cells (see Fig. 2).
Felidae: Also have a medulla with doughnut-shaped cells (Fig. 2). Can be distinguished from canidae in that the 'doughnuts' are larger and the medulla shows a sudden taper at the proximal end. Hairs usually exhibit distal banding.
Cervidae: Hairs have a very broad cellular medulla with a 'cobblestone' effect.
Bovidae: Cortex often has large ovoid bodies. Medulla is relatively narrow and can be absent, fragmental or continuous.
Equidae: Hairs often have transverse stria, and scanty pigment irregularly distributed.
Leporidae: Hairs have a multi serial ladder-type medulla and dog bone-shaped cross-sections.

As a supplement to macroscopic and microscopic animal hair identification, DNA and electrophoretic methods of species identification have recently been developed. It is anticipated that, in the future, such methods will gain widespread use in forensic laboratories.

Human Hair Identification

With human hair, the following information can usually be determined:

Body area of origin

Scalp hairs (Fig. 1) tend to be long with fairly constant diameters and cut tips. They can show evidence of treatment. Pubic hairs (**Fig. 3**) are generally kinky with wide variations in diameter. Their tips are frequently rounded or frayed from rubbing on clothing. Most beard hairs (**Fig. 4**) are broad in diameter, with triangular cross-sections. Eyebrow, eyelash and nasal hairs (**Fig. 5**) tend to be short and stubby, tapering to an abrupt point. Mascara or mucus may be adhering. Axillary hairs resemble pubic hairs except that they are less kinky and present a bleached appearance distally. The remaining hairs on the human body are generally indistinguishable from each other and are grouped together as general body hairs. It should be remembered that hairs on the periphery of various body regions will show combination characteristics, thereby making them difficult to identify (e.g. sideburns will show some characteristics of scalp hair and some of beard hair).

Method of removal

If a root sheath (**Fig. 6**) is present and/or the hair is in the anagen growth phase (**Fig. 7**), an examiner can conclude that the hair is indicative of forcible removal. (Forcible removal includes vigorous brushing or combing and inadvertent snagging.) A telogen root (**Fig. 8**), without a sheath, indicates that the hair fell out naturally. If the root is not present, an even break with regular edges indicates that it was cut off, and an irregular break generally means that the hair was broken off.

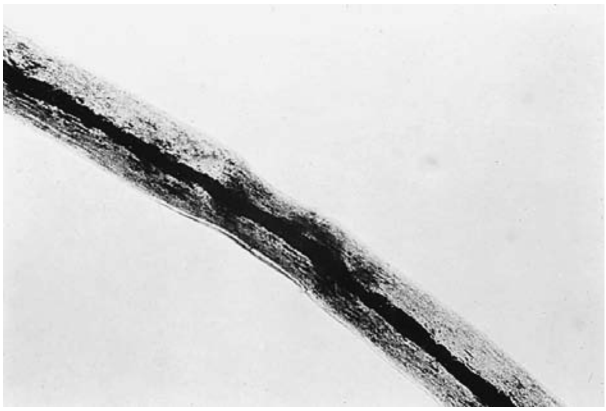


Figure 3 A human pubic hair (100 ×).

Racial origin

As a result of interbreeding there are few racially pure people. Accordingly, a forensic scientist can only give an indication of the racial origin of a hair. When stating racial origin, it should be pointed out that whereas a hair may show indications of a particular race, the other features of a person may or may not. Caucasian hairs generally have variable, nonblack color; fine to medium, fairly evenly distributed pigment granules; thin to medium cuticles; and only slight variations in diameter along the shaft. Negroid hairs are generally black or dark brown with medium

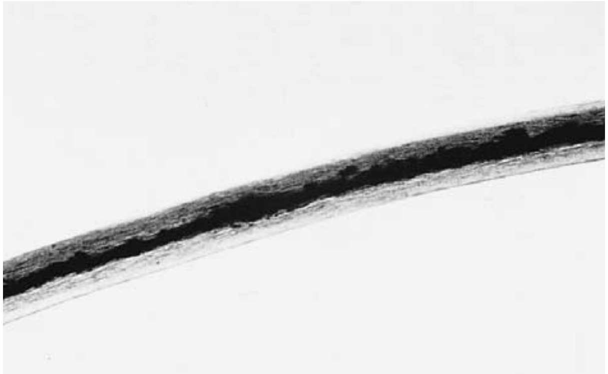


Figure 4 A human beard hair (100 ×).

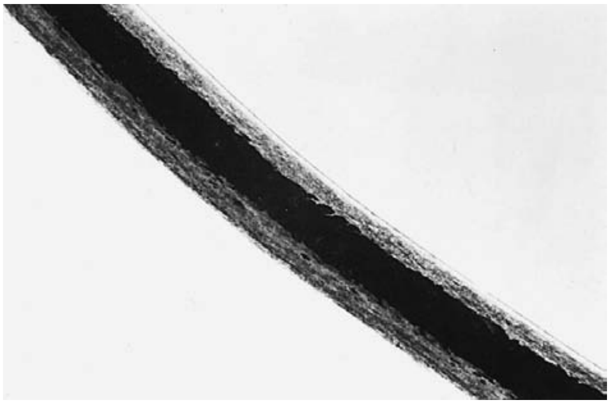


Figure 5 A human eyebrow hair (100 ×).

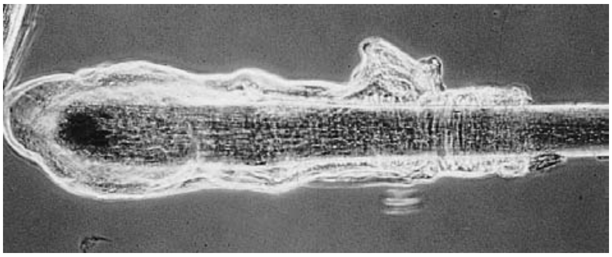


Figure 6 A root sheath attached to a human scalp hair telogen root.

Figure 7 A human scalp hair with an anagen root.

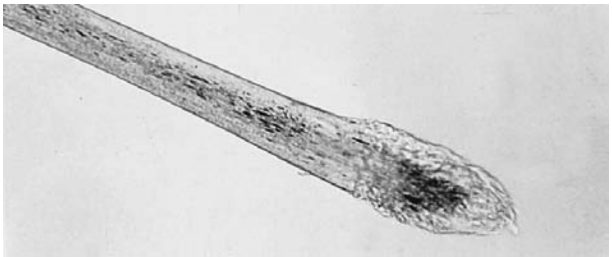


Figure 8 A human scalp hair with a telogen root.

to coarse pigment granules; clumped pigment; medium to thick cuticles; flat cross-sectional shapes and apparent wide variations in diameter along the shaft. Mongoloid hairs are usually black or dark brown with high pigment density; coarse pigment granules; thick cuticles; round cross-sectional shapes; and constant diameter along the shaft.

Cosmetic treatment

Microscopical examination An abrupt color change to a very light color indicates that the hair has been bleached.

Dyed hairs also show abrupt color changes and can have unnatural microscopic colors. The presence of color in the cuticle provides another indication of dyeing. Some dyes produce very subtle color changes. If an examiner is in doubt as to whether or not a hair has been dyed, a microscopical examination with incident fluorescence illumination can be conducted. At certain excitation wavelengths, an abrupt difference in fluorescent characteristics can often be detected between the dyed area of a hair and the undyed area. Knowledge of hair growth rates combined with measurement of the untreated portion of the hair can provide an estimate of the time since bleaching or dyeing. A curly appearance accompanied by constrictions in the shaft (and often by a damaged cuticle) is indicative of permanent waving. Examination of the hair under a scanning electron microscope can sometimes reveal thickly applied hair sprays.

Methylene blue infiltration test for bleached and permanent waved hair Place a hair, cleaned by washing in water and ethanol, in a shallow dish containing a solution of 5% NH_4OH in 25% by volume ethanol. Cover and allow to stand for about 10 min. Then transfer the hair to a 50% ethanol solution for another 10 min. After several washings with 95% ethanol, thoroughly air dry the hair and place it in a shallow dish. Cover the hair with a 1% solution of methylene blue stain and allow to stand for 5 min. Then rinse the hair with distilled water, wash with 95% ethanol, air dry the hair and mount it on a microscope slide using a mounting medium of refractive index of about 1.52 and a cover slip.

Under the microscope, bleached hair can then be detected as a penetration of the blue stain through the cuticle into the cortex. The amount of penetration of the dye and the depth of blue color produced will be proportional to the amount of bleaching of the hair.

Permanently waved hair can also be detected with this technique. In this case, the stain penetrates, but not to the extent seen with bleached hair. Usually there is penetration past the cuticle and slightly into the cortex. Instead of deep blue, the hair appears greenish blue.

Untreated hairs will be unaffected by the stain and will appear normal. Damaged hair will often allow penetration of the dye but the hair will not be stained with any consistency.

The methylene blue can be removed from the hair by demounting the hair, removing the mounting medium from the hair by rinsing in toluene, and placing the hair in a 10% acetic acid solution heated to almost boiling for about 6 min.

Another promising procedure uses staining with Rhodamine B in combination with fluorescence microscopy. It is claimed that this procedure gives a higher percentage of correct determinations of bleached and permanently waved hairs than does methylene blue staining.

Detection of semipermanent hair dyes Semipermanent hair dyes may be detected on hair by means of thin layer chromatography as used to detect dyes on fibers.

Detection of oxidative dyes and other hair care products Gas chromatography (GC) and gas chromatography–mass spectrometry (GC–MS) have been used to detect various hair-care products. Residues

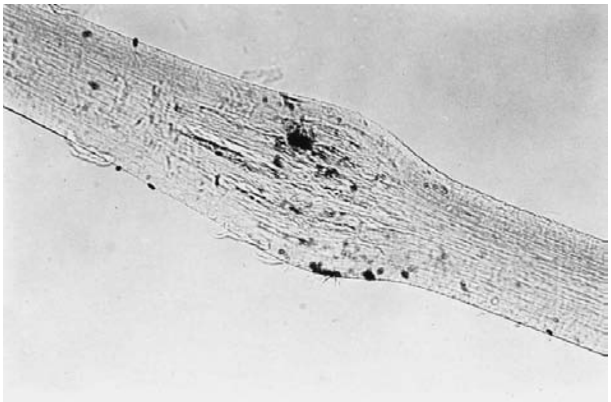


Figure 9 A crushed human scalp hair.

from 66 of 102 brands of hair care products (such as hair tonics, hair creams and waveset lotions) could be detected on hairs by means of GC and GC–MS analysis subsequent to ether extraction. In a later study with similar methodology, residues of four of twelve brands of hair spray and two of eleven brands of hair growth promoter were detected. Combination of the GC and GC–MS method with the thin layer chromatography method was found to increase discrimination in detection of oxidative hair dyes. However, identification of brand and type of hair-care products requires care because some components of oxidative hair dyes overlap with those of hair creams, waveset lotions, and hair rinses.

Damage

The shaft of a cut hair ends abruptly as opposed to the natural taper of uncut hair. If a hair has been recently cut, the margins are usually sharp and well-defined. With the passage of time, these margins become progressively more rounded off until the entire tip appears rounded.

Burned hair has a brittle charred appearance with large round vacuoles at the point of burning. Extent and type of damage depends on the temperature.

Crushed hair appears as in **Fig. 9**.

The tunneling of hair by fungal hyphae can produce distinctive transverse lines. Such damage can demonstrate that a hair has been exposed to fungi. This typically occurs with buried bodies. Another effect often seen in hair from people who have been dead for some time is termed 'postmortem root banding', whereby an opaque band about 0.5 mm above the root bulb can be observed with transmitted light microscopic examination.

Sex

At present, the sex of a hair donor can generally only be determined if a root sheath is present. The best methods for sex determination from hair root sheaths use forensic DNA analysis. In laboratories without access to forensic DNA analysis, the following procedure can be used.

After soaking the root and sheath in physiological saline for 3 h, immerse the hair in 25% v/v acetic acid for 1–2 min. After removing the excess acid and allowing the hair to air dry for 10 s, prepare cell 'tracks' by rubbing the hair bulb or root sheath in a zig-zag fashion within the circle on a specially prepared glass microscope slide. (Preparation involves inscribing a 1 cm diameter circle with a diamond pencil and cleaning the slide with alcohol.) Allow the preparation to air dry for at least 10 min, then cover the area of the circle with a 0.5%, w/v, aqueous quinacrine dihydrochloride stain. After 3 min, wash the stain with a gentle stream of tap water and immerse the slide in a beaker of water to destain for 6 min. To differentiate the cells, add McIlvaine's buffer at pH 5.5 (43.1 ml of 0.1 M citric acid and 56.9 ml of 0.2 M Na_2HPO_4), cover the preparation with a coverslip and let stand for 5 min. Read the slides at 1000 × on an incident light fluorescence microscope using blue light excitation and a 100 × objective equipped with an iris diaphragm. The Y-chromosome appears as a brightly fluorescing spot 0.5–1 mm in diameter. Less than 10% of the cells having such spots indicates that the hair is female; more than 50% spots indicates a male. It is advisable to confirm these determinations by an aceto-orcein staining procedure for sex chromatin in which the cells are stained with a cooled and filtered solution of 0.5 g synthetic orcein in 22 ml of hot (80–85°C) acetic acid and 27 ml distilled water. Examination is conducted at about 450 × under transmitted bright field illumination. In this test, staining of sex chromatin in more than 40% of the cells indicates a female and staining of sex chromatin in less than 10% of cells indicates a male.

A variation on the fluorescence procedure has been suggested whereby the same preparation is used for both X and Y chromosomes. Staining is conducted at pH 5.5 for Y chromosomes and at pH 3.0 for X chromosomes. The $Y-X$ score is then calculated, a score of 22 or more indicating a male and a score of less than 11 indicates a female.

Age

It is not generally possible to determine the age of an individual from which a hair originated. As an investigative aid, an experienced hair examiner can occasionally determine if a hair is from a very old or very young person. This is based on examination of the following characteristics which tend to increase with

age: medullary index, size of pigment granules, streakiness of pigment distribution, darkness of color, amount of cortical fusi, and size of unpigmented area just above the root.

See also: **Analytical Techniques**: Microscopy. **Fibres**: Identification and Comparison. **Hair**: Overview; Hair Transfer, Persistence and Recovery; Comparison: Microscopic; Comparison: Other; Deoxyribonucleic Acid Typing.

Further Reading

Appleyard HM (1960) *Guide to the Identification of Animal Fibres*. Leeds: Wool Industries Research Association.

Bruner H and Coman B (1974) *The Identification of Mammalian Hair*. Melbourne: Inkata Press.

Clement JL, Hegege R, Le Pareux A, Conet J and Gastaldi C (1981) New concepts about hair identification revealed by electron microscope studies. *Journal of Forensic Sciences* **36**:447–458.

Crocker EJ (1998) A new technique for the rapid simultaneous examination of medullae and cuticular patterns of hairs. *Microscope* **46**:169–173.

DeGaetano DH, Kempton J and Rowe WF (1992) Fungal tunneling of hair from a buried body. *Journal of Forensic Sciences* **37**:1048–1054.

Fujita Y, Nakayama M, Kanbara K *et al* (1989) Forensic chemical study on human hair. Part II: Identification of brand of women's hair care products by remaining components on human hair. *Eisei Kagaku* **35**:37–48.

Fujita Y, Yakamoto N, Nakayama M *et al* (1989) Forensic chemical study on human hair. Part III: Identification of oxidative dyes and presumption of the brands from remaining components on human hair. *Eisei Kagaku* **35**:444–453.

Fujita Y, Mitsuo MN and Satoh T (1990) Forensic chemical study on human hair. Part IV: Identification of brand of hair sprays and hair growth promoters by components remaining on human hair and example of forensic hair examination. *Eisei Kagaku* **36**:211–218.

Hicks J (1977) *Microscopy of Hairs: A Practical Guide and Manual*. Washington, DC: Federal Bureau of Investigation.

Luell E and Archer V (1964) Hair medulla variation with age in human males. *American Journal of Physical Anthropology* **22**:107–109.

Mayer HV (1952) The hair of California mammals with keys to the dorsal guard hair of California mammals. *American Midland Naturalist* **48**:480–512.

Moore TD, Spence LE and Dugnelle CE (1974) *Identification of the Dorsal Guard Hairs of Some Mammals of Wyoming*. Bulletin 14: Wyoming Game and Fish Department.

Mudd JL (1984) The determination of sex from forcibly removed hairs. *Journal of Forensic Science* **29**:1072–1080.

Palenik S (1983) Light microscopy of medullary microstructure in hair identification. *Microscope* **31**:129–137.

Petraco N (1981) A modified technique for the cross sectioning of hairs and fibers. *Journal of Police Science and Administration* **9**:448–450.

Petraco N (1987) A microscopical method to aid in the identification of animal hair. *Microscope* **35**:83–92.

Petraco N, Fraas C, Calley F and DeForest P (1988) The morphology and evidential significance of hair roots. *Journal of Forensic Science* **33**:68–76.

Robertson J (1999) *Forensic Examination of Hair*. London: Taylor & Francis.

Rosen S and Kerley E (1971) An epoxy method of embedding hair for histological sectioning. *Journal of Forensic Sciences* **16**:236–240.

Stegerda M (1941) Change in hair color with age. *Journal of Heredity* **32**:402–403.

Stocklassa B (1987) Fluorescence microscopy of single hair strands for detection of cosmetic treatments. *11th International Association of Forensic Sciences*. Triennial Meeting, Vancouver.

Appendix: Characteristics for Hair Identification and Comparison

The following characteristics were developed in 1998 by a subcommittee of TWGMAT (Technical Working Group for Materials Analysis Technology) and represent an updating of the list of characteristics originally developed in the early 1980s by the Committee on Forensic Hair Comparison.

- Hue: colorless; blonde; red; light brown; medium brown; dark brown; black; other
- Pigment density: absent; light; medium; heavy; opaque
- Pigment distribution: uniform; peripheral; one-sided; random; central/medial; pigment in cuticle; banded
- Pigment aggregation: streaked; clumped; patchy
- Pigment aggregate size: large; medium; small
- Pigment shape: round; oval; oblong; other
- Pigment size: large; medium; small
- Artificial treatment: dyed; bleached; other
- Length variation
- Curvature
- Form: straight; wavy; curly; twist; tightly coiled; crimp
- Diameter variation or range
- Cross-sectional shape: round; oval; triangular; flattened; other
- Other shaft configurations: buckling; convoluting; shouldering; undulating; splitting; regular; other
- Cuticular thickness: thin; medium; thick
- Outer cuticle margin: flattened/smooth; serrated; cracked; looped; other

- Inner cuticle margin: distinct; indistinct
- Appearance: color; texture; pigmentation; treatment; damage; other
- Cortex: cellular texture: coarse; uniform
  ovoid bodies: size; shape; color; distribution; abundance;
  cortical fusi: size; shape; distribution; abundance;
- Medulla: continuous; discontinuous; fragmented; opaque or translucent; relative width; amorphous or patterned; other; absent
- Proximal ends: root present: bulbous/telogen; catagen; anagen; sheathed; follicular tag; postmortem banding
  root absent: severed; decomposed; crushed; other
- Distal ends: tapered tips (uncut); rounded;
  cut: square; angle; frayed; split; crushed; broken; singed
  other
- Alterations other than color: environmental damage; mechanical damage; other
- Artifacts: lice; mold/fungus; bite marks; debris; blood; other
- Abnormalities: pili annulati; trichoschisis; monilethrix; trichorrhexis nodosa; trichorrhexis invaginati; pili torti; trichonodosis; cartilage hair hypoplasia

Handwriting *see* **Document Analysis**: Handwriting.

HEALTH AND SAFETY

Including Risk Assessment

D Schudel, J E McLaughlin and
C M Selavka, Massachusetts State Police Crime Laboratory, Sudbury, MA, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0563

Introduction

Health and safety are part of our professional lives. Forensic scientists are exposed to numerous and varied health risks on a daily basis, from the controlled setting of a laboratory, to the unpredictable nature of a crime scene. Health and safety measures are adopted to reduce the chance of an accident, but at some level it is the responsibility of each of us to check our working habits every day and address safety concerns as they appear. A good understanding of how accidents occur, and how they are avoided, not only reduces the chance of an accident in the workplace but can affect safety in other areas of our lives. A safe kitchen or garage is as important as a safe laboratory. The safety of our family and friends is as important as that of our colleagues.

Good health and safety comes from the top. If management does not actively support a health and safety program, it is unlikely that employees will follow suit. A commitment to safety involves not only written policies but also the allocation of money for equipment, time for designated safety personnel to do their tasks, meetings to discuss safety issues, and development of a clear expectation of disciplinary action for continued safety violations. However, there are many reasons for management to take an active role in supporting a safety program. Good health and safety means less time off work from accidents, reduced liability risk, better quality control and reduced risk of the contamination of evidence.

It is not possible to review all of the health and safety issues in this short article. The aim is to give the reader an overview from which to work, and encourage further learning from available texts, courses and conferences.

What is an Accident?

An accident is an unplanned event that causes some kind of harm. In many cases, the harm is obvious,

such as an acid burn to the arm, or exposure to a disease. In other cases, the harm may be more obscure, such as financial loss, the destruction of property from fire, or a mental scar.

Hazards

All accidents are caused by a hazard. A hazard is anything that can cause harm. In a forensic laboratory setting, hazards include chemical, physical, biological and electrical sources. At a crime scene, there may be other hazards in addition to the above, such as falling through a hole in a floor, or attack by a family pet. When identifying hazards it will seem that most things can cause harm under certain circumstances. Water in a container poses little threat, but spilled on a tiled floor is a slip hazard. The hazards in our work environment need to be addressed to assess their severity and the likelihood that they will cause harm. The chance of a hazard causing harm is called 'risk'.

How Accidents Occur

It is common to think of an accident as a single event. If a person trips over a power extension cord lying on the floor and falls, breaking an arm, we often address that accident as a one-off event and blame whoever put the power cord on the floor that day. However, accidents are typically not single events. It is usually the combination of several factors that leads up to the accident. In the case of the power cord, consider some of the other factors:

- How long has the power cord been on the floor?
- How many people tripped on the power cord and said nothing?
- Why was a power cord in use? Why was there no power in the area being worked?
- Who put the power cord there and why did they do it? Were they trained in the safe use of power cords?
- Who was supervising this area and allowed the power cord to stay?
- Were signs or mats available to warn of a trip hazard or to cover the cord entirely?
- How good was the lighting in that area?

The accident has been analyzed in terms of its 'direct cause' and 'indirect causes'. The direct cause of the broken arm was the person tripping on the cord and falling. The indirect causes are listed above. There is a domino effect of indirect causes leading to the direct cause. If one domino (indirect cause) is removed, then the direct cause (the accident) may not occur. The identification and rectification of indirect causes is key to accident prevention.

Risk Assessment

An effective risk management program is dependent upon the ability of laboratory personnel to identify and prioritize the presence of risks, and to identify potential hazards before a work-related accident occurs. This will assist in the establishment of effective work practices, engineering controls and administrative controls, which all contribute to a safe working environment. The forensic laboratory and the crime scene have a myriad of potential risks, due in part to the laboratory environment and in part to the nature of evidence examination. In order to perform risk assessment, two factors should be considered:

1. Predict the possibility of an accident occurring.
2. Determine the extent of the consequences should an accident occur.

Near Misses

A near-miss accident or incident occurs when an employee nearly has an accident but is able to avoid it. Assessment of near-miss incidents will also assist in identifying job hazards before a work-related accident occurs. Communication of common safety experiences may lead to the development of fresh alternatives to unsafe habits. Safety managers must initiate a reporting process for near-miss incidents. Each near miss, injury or accident should be investigated to determine the specific cause and whether any actions are needed to prevent recurrence. Develop and test the solutions, implementing the most effective. Ultimately, evaluate the results. Regular review of safety records will aid in the identification of injury patterns or new areas of concern.

Response to Incidents

The proper response to a dangerous situation lies at the heart of many of the elements that make up the safety and health program. You want the staff instinctively to make safe choices, to maximize the impact of the response and minimize threat and damage. Delivery on this goal relies on: (1) the realism and frequency of training; (2) promptness and completeness of response to incidents and near misses; and (3) documentation and use of historical information related to the program.

There is a symbiotic relationship involved with these three elements of the health and safety program. As you determine the training needs during start-up of the program, it is important to consider incidents that have happened in the past. You may have to rely

on anecdotes in the absence of firm documentation. As you listen to the stories, search for the root causes and determine which portions of the safety manual might address the underlying unsafe practice, equipment or procedure.

You are likely to find common issues at the root of previous safety incidents and near misses. Tailor your training priorities so that – in addition to required annual focus on fire response, bloodborne pathogens and chemical hygiene – you train the staff on the appropriate response and practices needed to avoid recurrence of historically prevalent problems. When a problem or near miss occurs, insure that you document it promptly and fully, and compile a running list of these incidents each quarter. Post the incidents so that staff can be informed of issues, and schedule training around the recurring problems. Alter the safety program to better prepare for and preclude recurrence of these problems, and alert supervisors to be particularly vigilant for root causes of these common issues during monthly audits.

It is a management truism (graphically depicted in the well-known Paretto chart) that 80% of significant issues of quality or safety can be traced to a single type of causative stimulus. If you identify and focus effort on eliminating this root stimulus, 80% of the remaining problems may be traced to another single causative stimulus. In any human system, you can never remove all the problems, but by systematically eliminating root safety concerns, the overall number of safety concerns and unsafe acts is quickly reduced. Systematic approaches and vigilance work together to provide a reasonable chance for success in the health and safety program.

The Facility

The setup of a laboratory is established when the facility is first occupied for use. Changes in technology and work tasks throughout the years may make the layout inadequate for current applications. Often we adapt to a laboratory setting that is no longer appropriate to contemporary usage. Assessment can include the design of workstations, scientific products, optical systems and displays to enhance the health and safety of employees. For example, consider relocating frequently used items to a more convenient and accessible area, thus reducing laborious reaching and bending. Examine laboratory office and bench chairs. Do these chairs provide the analyst with ergonomic comfort consistent with the job function? Improving comfort for the examiner who performs microscopic analysis can reduce eye, back and forearm strain. The forensic analyst often spends long hours at the laboratory bench. Appropriate height of the workbench surface will promote neutral postures, thus reducing development of chronic strain and injury.

Storage

Safe and secure evidence storage space that does not pose a health hazard from contaminated clothing, etc. is a taxing problem for many forensic laboratories. Often, large volumes of evidence must be stored prior to examination. Assessment of the evidence storage area should include consideration of appropriate storage height. Heavy items should be stored at heights that do not require overhead reaching or increased bending. Ladders or stools should be available to retrieve evidence stored above chest height. Employees should be trained in proper lifting techniques to reduce the possibility of back strain and injury.

Specific attention must be paid to the proper storage and handling of compressed gases. Gas cylinders are 'rockets in waiting', and must be secured and appropriately labeled at all times.

Obstructions

Attention to proper housekeeping and maintenance may reduce the risk of trips and falls that occur as a result of aisle clutter. In general, an obstructed aisle will not lead to a severe accident. Employees may suffer a strain or bruise as a result of tripping over an obstruction. However, the consequences rise dramatically if, in the event of an emergency, an obstructed aisle prevents or delays an employee from leaving the facility.

Electrical

Many forensic laboratories have a large inventory of computer-driven instrumentation, including spectrophotometers, gas chromatographs with various detectors, scanning electron microscopes and DNA sequencers. In addition to the risk of shock, short circuits can produce electrical arcs, leading to injury or fire. Proactive intervention begins with mechanical measures. Safe installation of electrical equipment by qualified personnel, combined with insulation proper to the voltage, will reduce shock hazards. Risk assessment should include proper maintenance of equipment in order to reduce the risk of deterioration creating an unsafe condition. Equipment should be located in a space sufficient to allow the heat produced to ventilate properly. Electrical cords must be appropriately secured, reducing the chance of trips or accidental disconnection. Personnel should be trained

to inspect cords for signs of wear routinely and replace as needed.

Body Fluids

The forensic scientist routinely examines clothing and evidence contaminated with body fluids. The risk of exposure is a frequently occurring event with severe consequences. Personal protective equipment reduces the potential for an exposure to these fluids. Management must provide employees with personal protective equipment that is comfortable and appropriate to the task. Comfort will contribute to improved compliance. Protective measures against infectious agents should be directed against the routes of transmission and exposure. Gloves provide a protective barrier while handling evidence. Mucous membranes should be protected with a facemask in combination with goggles or full-face shield whenever working with liquid body fluids or their extracts. These measures must also be reinforced by a general attention to laboratory cleanliness and hygiene. Employees must regularly inspect and decontaminate reusable receptacles, such as buckets, centrifuges, cans and carts, that may be contaminated with infectious materials. Cleaning and decontamination of equipment work surfaces and utensils are an effective means of minimizing accidental transmission of diseases.

Sharps

Reducing the risk of puncture or cuts involves proper training in combination with adequate protective equipment. The forensic chemist may perform an examination of used hypodermic syringes for the presence of controlled substances. Cocaine and heroin intravenous users are high-risk populations for hepatitis, HIV infection and other infectious diseases. Proper training involves containing the syringe and needle in a leak-proof, puncture-resistant container before submission for analysis. Gloves should be donned prior to handling the syringe. Skin punctures usually occur as a result of recapping the needle; this, therefore, should not be practiced. After examination the syringe should be returned, uncapped, to the container. The container is then sealed and returned to evidence. Accidents involving broken glass should be handled with mechanical means, such as forceps or a dustpan and brush, thus reducing the possibility of cuts. If the broken glass object contained, or was contaminated with, blood or other potentially infectious material, remember to disinfect the clean-up utensils after use.

Chemicals

In addition to body fluids, the forensic scientist will routinely handle chemicals in various analyses. Eye protection in the form of goggles or full-face shield should be considered when the possibility of liquid splashes or spills exists. Assessment of risks should include response to spills. A convenient location for maintenance of the Material Safety Data Sheet (MSDS) library should be established. Access to MSDS, safety manuals and other safety documentation may be assured through computers located throughout the laboratory. Each laboratory employee should be well informed as to the location and how to interpret the information of an MSDS.

Spill kits should be readily available to employees who have been trained in proper clean-up. Spill response should include immediate notification to personnel that a spill has occurred, donning protective equipment prior to containment, clean-up and proper disposal. Chemical spill training will reduce the possibility of hazardous chemical exposure to employees.

Hoods

Routine inspection of chemical and biological hoods will protect employees from exposure to hazardous vapors and infectious materials. Hoods containing a filter system should have scheduled filter replacement appropriate to the manufacturer's recommendation. Face velocity flow tests should be conducted routinely to assure proper hood operation.

Scene Work

Hazards in the laboratory occur in a controlled setting. In many ways, it is the often unpredictable nature of the scene that poses unforeseen hazards and far higher risks for the forensic scientist. There are four basic scene scenarios: crime scene, autopsy, fire scene and clandestine laboratory.

Crime scene

Response to crime scenes requires forensic personnel to be adequately trained in health and safety issues. Each member of the response team must make observations and respond appropriately. The crime scene can be in places as different as the bilge area of a ship or the side of a highway. Physical hazards must be assessed before any work is done. The scientist must ask: 'Can I be hit by a vehicle? Is the scene a confined space? Do I have to worry about getting too cold or too hot? Can I fall?' The assessment of hazards will depend on the scene and should be discussed with all parties involved.

Biological hazards are a major concern in any nonlaboratory setting. Blood may be fresh and unconfined. The risk of splash on the face or skin is high. The floor may even be slippery with blood. There may be other hazards not encountered in the laboratory. In an outdoor scene, animal feces can harbor fungal and bacterial infections, such as 'valley fever' from concentrations of bird droppings. Deer ticks can spread Lyme disease. Some spider bites cause serious illness or death. Use a rapid assessment of attendant risks to guide choices for clothing, assisted ventilation and individual roles of available personnel.

Autopsy

The autopsy occurs in a controlled setting with safety precautions generally in place. However, attendants at autopsy need to be aware of the risk of infection, not only from contact with blood but also from inhaled airborne particles. Saws can generate particulates that carry infectious fomites. Tuberculosis is spread by airborne nuclei and is a risk at autopsy or when moving a body at a crime scene, when airborne particles may be forced from the lungs. Any work with the brain or central nervous system may carry the increased risk of Creutzfeldt–Jakob disease (CJD). The causes of this disease, protein-based prions, are highly resistant to normal sterilization techniques.

Fire scene

Fire scenes carry the normal hazards of a typical crime scene, but the added danger of what is often a highly uncontrolled environment. Firefighters may be aware of these dangers, but the forensic scientist with little fire scene experience may not. Before entering the scene, check that the electricity has been made safe (usually it is switched off at the source) and that other utilities (gas and sewerage) will not pose a hazard. Other hazards include holes in the floor, danger of a structural collapse and wires poking out at eye level. For those who attend many fire scenes, unprotected breathing of airborne gases and dusts, such as charcoal dust, may cause long-term health problems.

Clandestine laboratories

The major hazard at a clandestine laboratory is the use of flammable liquids by miscreants who are not fully aware of their dangers. The atmosphere in the laboratory may be a fire or explosion waiting to be ignited. At dangerous exposure levels, inhalation of the vapours may result in disorientation, possible loss of consciousness, or caustic insult. Ignitable catalytic metals may also be present. The drugs, reactants and byproducts may be present in the air, or on many surfaces, creating risks for contact or inhalation intoxication. Other possible hazards may exist, such as deliberate traps designed to hurt or kill unwanted intruders. Investigators should be adequately trained in the use of respiratory protection, chemical-resistant suits and appropriate gloves. Each laboratory is unique and may require confined space considerations due to poor ventilation, limited egress or hazardous atmosphere.

Sources of Information, Guidelines and Law

The standards to be met in safety program design and application must meet the requirements of many different agencies. In addition, the forensic laboratory exists in a jumble of regulations, policies, legislation, case law and voluntary guidelines. These must be synthesized to make the local application practicable for the laboratory employee, yet bulletproof to scrutiny from an agent of one of the various bodies named above. As a rule, any single approach that we could recommend for laboratories would constitute 'one size fits none', so we will not attempt it. On the other hand, certain general principles for optimizing to meet local requirements apply to all forensic laboratories.

All laboratory procedures, policies and practices must meet applicable national, provincial, state, municipal and local legislation. There must be no gray area about which laws apply: consult your agency counsel, and those at the agencies higher on the food chain, to insure that you are aware of them all. Remember also that formally promulgated rules and regulations of the government have the legal weight and enforceability of law. Therefore, compliance is not optional for these rules and regulations.

You should ask other forensic laboratory administrators in the same or similar situations which laws they feel they must meet. Insure that your laboratory is compliant if your situation cannot be legally differentiated from theirs. Research case law that has developed in your jurisdiction as a result of worker exposures, injuries or fatalities. Understand the cause of the actions and the specific wrongful acts of management addressed through the court's findings. A call to attorneys involved, public or employee advocacy groups and union representatives for your staff (if applicable) will help round out the story. There is no better way to avoid the problems of the past than to ask those involved on both sides. From these stories you must learn which legal interpretations have been wrong in the past, or specific actions and policies required to meet legal intent.

Local policies related to safety from your laboratory's host department and agency should be reviewed as well. However, these do not bear the weight of law, and often have not necessarily been subject to the same rigorous deliberation before implementation. This makes them more negotiable should you find incongruence with legislative, case law or regulatory requirements. When you find incompatibility, you must comply with laws and regulations. Insure that your program complies with all applicable portions, and document your reasoning if you should find it necessary to not comply with any portion of the local policy.

In industrialized nations, safety and health have not been in the purview of 'voluntary processes' for some time. Owing to employer ignorance, disregard and past abuses, most forensic laboratories must be compliant with strict legal requirements as outlined above. However, voluntary programs may provide guidance to the laboratory in fleshing out the safety and health program, to improve it beyond the level of the legal standard.

Virtually all of the accreditation organizations in forensic science – including the National Association for Measurements and Standards (NAMAS, UK); National Association of Testing Authorities (NATA, Australia); Standards Council of Canada (SCC), American Society of Crime Laboratory Directors/ Laboratory Accreditation Board (ASCLD/LAB) and American Board of Forensic Toxicologists (ABFT) – incorporate essential elements in their accreditation criteria governing laboratory safety and health. Specific guidance for meeting these essential criteria may be obtained in the documents provided to applicants by the organizations. Updates and specific changes to the interpretation of these criteria are provided in the organizations' newsletters and proceedings.

In addition, organizations such as the International Organization for Standardization (ISO), British Standards Institution (BSI) and American Society for Testing and Materials (ASTM) may come to mind when considering the availability of consensus standards for safety and health in the workplace. You may consult model documents from international and national trade organizations specializing in scientific workplace safety and health, such as the National Safety Council. In addition, the American Chemical Society and other national scientific organizations provide training programs for laboratory safety workers and managers, as well as implementation advice and training in risk assessment. Most of these programs are designed to insure compliance with 'current Good Laboratory Practice' (cGLP) standards, upon which rests the commercial laboratory's ability to compete in the international marketplace.

Forensic laboratories are not usually considered in this competitive sense, but the forces of privatization at work in the United Kingdom and Australia may be harbingers for such conversion of forensic infrastructure in other countries. Therefore, laboratory managers and safety and health officials should at least be informed of the availability of such standards and training programs from commercial sources, and the role they serve in meeting competitive industrial requirements for free trade.

Policy Content

The manual of safety and health for a laboratory often matches the personality of its leadership and safety advocates. Some create massive safety tomes containing details ranging from imperatives to the trivial. Others take the short road and create policy manuals that fulfill legal requirements and provide reasonable but brief guidance on meeting a stated policy for health and safety maintenance without undue prescription. Your policy will be a personalized work, but recognize that the greater the level of detail, the faster the obsolescence and greater the requirement for routine scrutiny and update. But you must have reasonable detail so that staff members can understand and comply with the law.

As you develop your manual, here are some section headers routinely included in laboratory safety and health policy manuals. Modification will be made to fit your own situation, as 'one size fits none' is not our intent:

- *Employers' statement of intent*: A brief summary of the necessity for, and management support of, the policies in the manual.
- *Powers of the safety officer*: Assignment of responsibilities and authority to the safety officer and committee members for issuing recommendations and monitoring compliance.
- *Employee responsibilities*: Assignment of responsibilities of employees in fulfilling elements of the policies and specific reporting requirements.
- *Program for new employees*: Specific elements of training and familiarization which must be completed before release of new employees to perform laboratory work.
- *General guidelines*: Overall program of safety, often including personal protective devices, physical plant elements of safety and health, update and maintenance procedures, reporting mechanisms and procedures for requesting policy clarifications and changes.
- *Specific hazards*: Guidelines for insuring a safe and healthy laboratory environment when dealing with

problems unique to the laboratory design, type of work, type of evidence or as otherwise dictated (see review of forensic hazards above).

- *Material safety data sheets*: Instructions for accessing and using the information in these important documents. (Note: the documents themselves are not usually contained in the safety manual.)
- *Rewards and discipline*: Specific administrative outcomes to be expected for compliance with, or disregard for, the policies in the safety manual.
- *Record-keeping requirements*: Procedures for routine review, update, change and historical archiving of the manual.

Clearly, the laboratory will add other sections to the manual, and modify the content of each, as dictated by consideration of law, organizational requirements and the needs for continued improvement. Let history be your teacher.

Implementing Health and Safety Programs

To be effective, the program of health and safety must become integrated in the fiber of daily life at the laboratory. It is not enough to have excellent manuals on the shelf and records that demonstrate that employees have read and understand these manuals. Getting the staff to think of safety first in all procedures can become a test of wills. It is human nature to concentrate on the rush case, the broken instrument or the persistent pathologist's question. Should you expect true understanding and unerring compliance? In a word, no. But you can build a series of rewards and punishments that will provide reasonable motivation to all employees to think and act safely, and construct a paper trail of appropriate responses when the odd staff member continually refuses to comply.

Never forget that safety starts at the top. The laboratory director, managers and supervisors have the legal and moral responsibility for insuring the availability of a safe and healthy workplace for everyone. Therefore, the job descriptions, goals, objectives and performance measures for each of these staff members must specifically set forth this duty. For example, supervisors' objectives could include:

- Through his or her own action, demonstrates appropriate behavior to meet all requirements of the laboratory safety and health program.
- Through proactive measures, insures compliance among his or her staff with all requirements of the laboratory safety and health program.
- Performs routine safety audits of his or her own unit, and other units as assigned, and performs appropriate follow-up and reporting as required by the laboratory safety and health program.
- Appropriately rewards compliance, or performs disciplinary action related to noncompliance, with the laboratory safety and health program by subordinate staff members.

In addition, *all* employees should have at least the first of these sample objectives included in his or her documented duties. Performance against these objectives becomes a normal part of the process by which management evaluates employees. Be fair in this evaluation, and reward significant compliance and proactivity when they are observed. Positive reinforcement of proper action, and contributions to the improvement of the health and safety program, must be documented and rewarded. Make a point of catching staff when they do something right, even if the reward in such cases is only a quick note with a copy in the file.

A pattern of noncompliance should lead to corrective action, and continued noncompliance thereafter should be approached using progressive discipline. The hard part of this equation, as with any objectionable behavior, is determining when a behavior constitutes a pattern. The watchword is to be consistent, and document your actions.

In addition to the measures listed above, various strategies may assist the acceptance of the program by staff. Holding a safety fair or safety Olympics, posting safety-related information on a bulletin board, making safety discussions part of all continuing education programs, and generally keeping safety in front of everyone's nose as often as possible will help. However, the single most effective way to capture and keep the laboratory focused on safety is to regularly audit against established standards of practice. The adage 'What gets measured gets done' is never more true than with safety.

The standards for audits are the safety manual and its interpretive discussions. Audits should be performed monthly, using checklists for ease and consistency of evaluation, and should be performed by supervisory staff on their own and each others' units. Training for auditors should be included in skill development sessions with all supervisors. This training should also be structured to allow supervisors to evaluate and assess potential risks, using previous incidents and near-miss reports for guidance.

See also: **Accreditation of Forensic Science Laboratories**. **Administration of Forensic Science**: Organization of Laboratories. **Basic Principles of Forensic Science**. **Crime-scene Investigation and Examination**: Recovery of Human Remains. **Fire Investigation**: Fire-scene.

Further Reading

Baldwin D (ed) (1992) Safety and Environmental Training. ITP: Van Noystrum.

Forsberg K and Mansdorf SZ (eds) (1997) *Quick Selection Guide to Chemical Protective Clothing*, 3rd edn. ITP: Van Noystrum.

Furr (ed) (1995) *CRC Handbook of Laboratory Safety*, 4th edn. Boca Raton FL: CRC Press.

Goetsch DL (ed) (1996) *Occupational Safety and Health in the Age of High Technology*. New York: Prentice Hall.

Kletz T (ed) (1993) *Lessons from Disaster*. Gulf.

Kletz T (ed) (1993) *Management of Environmental Protection and Safety*. Gulf.

Lunn GL and Sansone EB (eds) (1994) *Destruction of Hazardous Chemicals in the Laboratory*. Chichester: Wiley.

Patnaik E (ed) (1992) *A Comprehensive Guide to the Hazardous Properties of Chemical Substances*. Chichester: Wiley.

Pipitone D (ed) *Safe Storage of Laboratory Chemicals*. Chichester: Wiley.

Sittig M (ed) (1991) *Handbook of Toxic and Hazardous Chemicals and Carcinogens*, 3rd edn. Noyes.

Wallace I (ed) (1996) *Developing Effective Safety Systems*. Gulf.

Height *see* **Anthropology**: Stature Estimation from the Skeleton.

Histopathology *see* **Pathology**: Histopathology.

HISTORY

Contents

Crime-scene Sciences
Fingerprint Sciences
Forensic Anthropology in the USA
Forensic Medicine
Forensic Sciences

Crime-scene Sciences

G C Knupfer, National Training Centre for Scientific Support to Crime Investigation, County Durham, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0566

Introduction

The discipline of crime scene investigation developed in Europe and elsewhere in the second half of the nineteenth century, where such figures as Alphonse Bertillion, an employee of the (French) Sûreté, began experimenting with the techniques of anthropometry (or bertillionage) – an elaborate system of recording a convicted criminal's body measurements for identification purposes. Perhaps more significantly, however, he began recording crime scenes by using photography, and developed techniques for casting tool marks to identify instruments used in the course of crime.

In contrast, a number of British enthusiasts, skeptical of Bertillion's identification techniques for convicts, were concentrating on the study of fingerprints as the definitive method of identification. In 1901, a fingerprint bureau was established in Scotland Yard, England, and in 1902 the first conviction was obtained using fingerprint evidence.

However, it was Dr Edmund Locard, a pioneer of scientific criminology based in Lyon, who, in the early twentieth century, developed his now famous and generally acknowledged law of forensic science: 'Every contact leaves a trace.' In other words, when two objects come into contact, there is a mutual transfer of material, no matter how slight. For example, when two fabrics touch, it is likely that there will be a cross-transfer of fibers from one to the other.

Crime scene examination, as a separate discipline or science, grew out of these three techniques: crime scene recording and photography; the search for, recording, and recovery of fingerprints; and the search for and recovery of forensic evidence. In some jurisdictions the discipline is known as crime scene investigation, and the practitioners variously as 'crime scene technicians', 'crime scene examiners', 'crime scene investigators', or 'scenes of crime officers'.

The pace and course of these developments varied from one country to another, as law enforcement agencies attempted to cater for their respective judicial systems and other national influences. In an article of this size it is difficult to reflect these many variations. The author will, therefore, concentrate on describing the development of the discipline in the United Kingdom as a model. It is reasonable to conclude that broadly similar systems will have developed and be found in many other countries throughout the world.

Development of the Role in the UK

The development of the modern British police service can be traced back to the formation of the Metropolitan Police (London) in 1829. In the ensuing years, police forces were established in the counties, towns, and boroughs throughout the United Kingdom, and, by 1856, every part of the British Isles was served by a police force. It is estimated that there were some 400 forces in existence at this time.

Forensic medicine

Forensic medicine has a long and celebrated history in the United Kingdom. As early as 1807, the University of Edinburgh had a chair in forensic medicine and the University of Glasgow created a similar post in 1839. In 1834, Alfred Taylor was appointed Professor of Medical Jurisprudence at Guy's Hospital, London, and 2 years later published his *English Elements of Medical Jurisprudence*. While medical evidence was given for the first time in a murder trial in 1859, it was not until 1900, that Doctors Luff, Wilcox, and Pepper, of St Mary's Hospital, Paddington, were retained as the first Home Office medical experts.

Perhaps the most well-known name of this era was that of Sir Bernard Spilsbury, a pathologist who, having followed in the footsteps of Luff, Wilcox, and Pepper at St Mary's Hospital, was retained by the Home Office and gave evidence in countless murder cases in the early decades of the twentieth century.

Fingerprint technology

The second half of the nineteenth century saw dramatic steps taken in the development of the science of fingerprint examination by such notable names as Sir William Herschell, Sir Francis Galton and Sir Edward Henry. Disillusioned with Bertillion's anthropometry, Scotland Yard created one of the world's first fingerprint bureaux in 1901. Developments were rapid. The first arrest and conviction for burglary based on fingerprint evidence occurred in 1902, and the first murder conviction based on fingerprint evidence occurred in 1905. A similar arrest and conviction for murder occurred in Chicago in the United States in 1911.

However, it is reasonable to conclude that not all British police forces moved forward at the pace of the Metropolitan Police. For example, the following extract is taken from a police instruction book dated 1923, issued to officers in one of the larger city forces:

> In cases of breaking into premises, finger impressions are often left on some parts of the building, or on some of the furniture, or other articles therein ... In all cases of this nature ... a special search must be made by the officer called to the scene of crime with a view to the discovery of such finger impressions, particularly upon glass, paper, or polished surfaces. In the event of any impressions being discovered they must be carefully protected until the arrival of a superior officer, when proper means will be taken to have the impressions compared with those of the Register of Habitual Criminals at London.

Nevertheless, it was to support these new techniques that police officers began to specialize in the search and recovery of fingerprints from crime scenes. While the first crime scene examiners were fingerprint experts in their own right, their successors developed a broader base, becoming skilled in scene photography and the search for and recovery of other evidence types, such as blood, fibers, and hair.

Forensic science

In the early part of the twentieth century, forensic expertise was provided by practitioners who were sought out by the police. These experts might have been analytical chemists and biologists working in hospitals or perhaps engineers and textile experts working for various industrial institutions. A good

example, for instance, was Robert Churchill, a celebrated London gunsmith, who gave evidence as an expert for the first time in a murder trial in 1910.

While the role of the 'Home Office pathologist' (see above) lives on to this day, it was realized at an early stage that the *ad hoc* hiring of forensic experts was an inappropriate and ineffective way of providing forensic science to the police service. As a consequence, in 1932, the first of a series of Home Office forensic laboratories was established in Nottingham. In the following years further laboratories were established at regional centers around the country, as well as in Scotland and Northern Ireland. Effectively, this forensic service was provided by the Home Office to police forces free of charge.

The Crime Scene Examiner/Investigator

It is within this context that the role of the crime scene examiner has developed, particularly in the second half of the twentieth century. The United Kingdom currently has some 52 police forces, all with access to fingerprint bureaux and laboratories and all employing crime scene examiners. In some forces, crime scene examiners are serving police officers, whereas in others, civilian employees undertake this role. There has been a concerted drive in recent years to return police officers to front-line policing roles, leaving crime scene examination to be undertaken by nonpolice personnel. While some experience and expertise has inevitably been lost in this process, the quality of many of the new examiners is very impressive. A number, for example, already hold postgraduate degrees in forensic science.

It would, however, be no understatement to say that the journey to the position in which we currently find ourselves has, on occasions, been far from smooth. The 1960s, 1970s, and early 1980s saw numerous miscarriages of justice coming to light where, in the absence of hard forensic evidence, courts had incorrectly relied on verbal admissions and confessions gained by investigators. Other examples, particularly in terrorist investigations, where forensic science and crime scene examination had clearly failed the judicial process, are also often quoted.

One such example is the case of the 'Yorkshire Ripper', a man who, between 1977 and 1982, murdered 13 women and attacked several more. He was finally arrested (albeit inadvertently by a uniformed police patrol) and convicted of multiple murder. In a subsequent review of the cases it was noted that, among other things, there was a distinct lack of any coordination of the forensic science input in the investigations. It has been suggested by some observers that, had evidence and information recovered from the crime scene been handled more effectively, the offender, Peter Sutcliffe, might have been arrested at an earlier stage. In the light of the many other difficulties experienced in the Sutcliffe case, not least in managing and analysing the sheer volume of information coming into the incident room, the UK's computerized major enquiry system (HOLMES) was born. This system has recently been upgraded and has served the police service very well over the intervening years.

Another notorious milestone was the case of Jeremy Bamber. In 1985, Bamber telephoned his local police force stating that he had received a distraught telephone call a few moments earlier from his father. Bamber Senior resided nearby in a large farmhouse with his wife, his daughter, and her two young children. Bamber Senior allegedly told Jeremy that his daughter, who had previously suffered from psychiatric problems, was going berserk in the house armed with a .22 caliber rifle. Bamber Senior then rang off. Jeremy Bamber went to the farmhouse together with several police officers, where they found that Bamber Senior, Mrs Bamber, their daughter, and her two children had all been shot dead. The scene suggested that Mr and Mrs Bamber and the two young children had all been murdered by the daughter, who had then lain down on a bed and committed suicide by shooting herself in the mouth.

The police handled the case throughout as one of four murders and a suicide. However, some time later, Jeremy Bamber's girlfriend contacted the police and reported that Jeremy had in fact murdered the entire family for the inheritance. When police returned to the scene, they found an abundance of evidence to support this revised scenario. For example, a sound moderator (silencer) suitable for the murder weapon was found in a cupboard. The silencer had traces of the daughter's blood inside it. Further, had the silencer been fitted to the murder weapon, the daughter could not have reached the trigger while the end of the barrel was in her mouth.

The police investigation of this incident was very heavily criticized and a number of fundamental recommendations resulted, outlining how the police should handle major crime scenes in the future. These recommendations were consolidated and enhanced in the early 1990s by the Association of Chief Police Officers (ACPO), who published a number of manuals and documents formulating policy on the management of major incident scenes. One such paper is known as MIRSAP (Major Investigation Incident Room Standardized Administrative Procedures). MIRSAP has recently been revised. It highlights the roles of the scientific support coordinator,

the crime scene manager, and the other specialists employed at the homicide or major incident scene. **Figure 1** illustrates the relationship between the senior investigating officer (SIO), the crime scene coordinator, and the crime scene manager. The crime scene (or scientific support) coordinator acts as the advisor to, and conduit between, the senior investigator (as well as individuals involved in other facets of the investigation) and members of the crime scene examination teams. A major investigation will have only one crime scene coordinator, but all the different scenes emerging in the course of the inquiry (e.g. victim, suspect, suspect's car, accomplice, etc.) will each warrant a separate crime scene manager. The object of the exercise here being to avoid cross-contamination from one scene to another. Essentially, the crime scene coordinator's role is 'hands off', while that of the crime scene managers and their staff is very much 'hands on'. This system, although relatively new, is now working effectively in the UK and cutting off many of the avenues of defense traditionally explored by some of the more forensically aware defense lawyers.

One of the other 'good practice' developments of recent years has been the separation of the roles of forensic examination and forensic search. There are countless tales of mistakes being made in investigations because crime scene examiners have not been correctly briefed. It is now argued that the role of crime scene examination is undoubtedly the role of the crime scene examiner (investigator), while that of searching the scene and its surroundings is that of the trained and qualified police searcher (known in the UK as a police search advisor, POLSA). An illustration of these different roles can be gained from a recent murder case in which a criminal lawyer murdered his wife and alleged it was the work of a third party. The murder had occurred in the marital home but, being forensically aware, the husband had removed or avoided the pitfall of leaving any forensic evidence around for the crime scene examination team to find. Upon interview, he was absolutely adamant that his wife's house keys had been taken by the offender. A detailed forensic *examination* of the house, lasting some 5 days, revealed absolutely nothing to implicate the husband (a strong suspect), or indeed anyone else. However, this process was followed by a forensic *search*. Appropriately trained (POLSA) officers spent several more days searching the house from top to bottom, including under the floors. Hidden behind a brick in the foundations, they found the deceased's house keys. It would be fair to conclude that the lawyer was convicted of the murder on this evidence alone.

One of the most striking developments of recent years has been the introduction of DNA technology

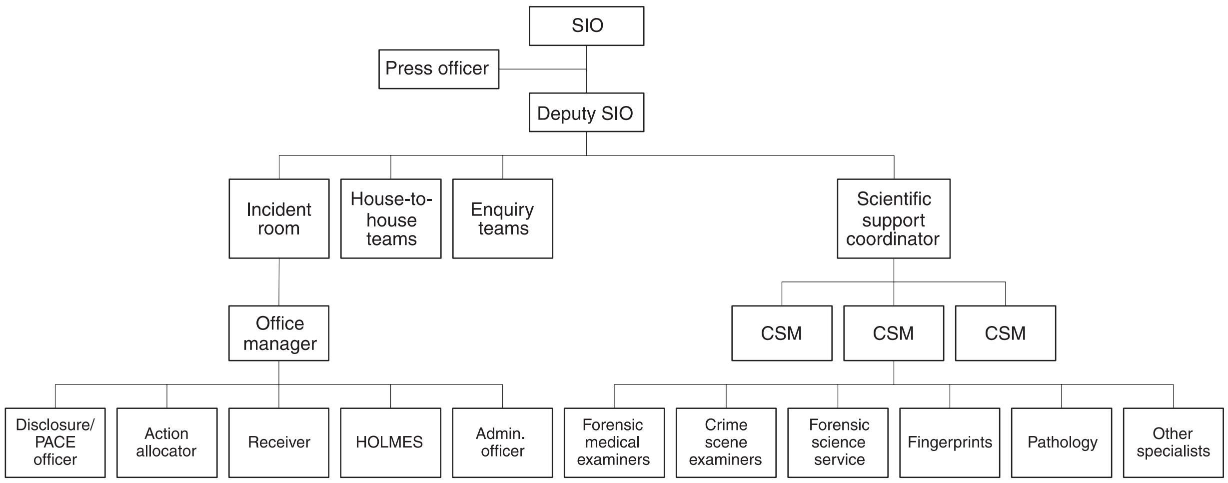


Figure 1 Typical major incident investigation structure.

to the field of crime scene investigation. This significant breakthrough came in 1987 with the trial of Colin Pitchfork, who, between 1983 and 1986, raped and murdered two young women in Leicestershire. The investigators sought the assistance of a biologist, Dr Alec Jeffreys, then working in the field of DNA research at Leicester University. Using this new technique, Jeffreys was not only able to eliminate an early prime suspect for the offences but, following a mass screen of 'eligible' males in the area of the murders, was able to identify Pitchfork as the offender. Pitchfork was subsequently convicted of both offences and sentenced to life imprisonment. This case placed DNA technology firmly on the investigative map.

The technology has moved forward at a breathtaking pace, from the multilocus probe (MLP) technique of the early 1980s, to the current second-generation multiplex (SGM) procedures. Third-generation multiplex (TGM) is planned for the year 2001. It would be no exaggeration to say that DNA profiling is now used to great effect on a daily basis in the majority of crime scene investigations, from homicide to auto crime.

The effectiveness of this technology has been aided, to a significant extent, by the passing of legislation in 1995 enabling the police to take buccal (mouth) swabs from individuals charged and convicted of the vast majority of criminal offenses, notably crimes of violence, sexual offenses, and offenses of dishonesty. These samples are then analyzed and the profiles lodged on a database maintained by the (UK) Forensic Science Service, on behalf of the police service. DNA profiles recovered from crime scene investigations are routinely checked against the database, with increasingly impressive results.

Training

In the late 1980s, a number of reviews were commissioned into the workings of the criminal justice system, and particularly the effectiveness of forensic science in the investigation process. Several recommendations were made with a view to improving the recovery of forensic evidence at the crime scene. One such recommendation addressed the issue of training and advocated the establishment of a national training center for scientific support to crime investigation to train crime scene examiners and fingerprint experts. The center was duly established in County Durham in 1990. The center serves the whole of the United Kingdom with the exception of the Metropolitan Police (London). Being by far and away the largest force in Britain, the 'Met' elected to maintain its own training establishment. While attendance on these 'national' courses is not mandatory, it is certainly recommended good practice and the majority of forces now send their personnel to the center. The courses are also open to overseas students, who are attending in increasing numbers.

The core course provided by the center are: initial, development and refresher crime scene examiner courses, and initial, intermediate, and advanced fingerprint courses. Other specialist courses provided include: crime scene management, scientific support (crime scene) coordination, fire investigation, facial identification, fingerprint evidence recovery and recording techniques, dealing with major disaster scenes and forensic medical examiner courses.

A number of higher educational qualifications are now available to scientific support staff. The national training centre has formal links with the University of Durham and jointly offers a 2 year part-time Diploma in Crime Scene Examination. A similar Diploma in Fingerprint Examination is planned for the year 2000. The Metropolitan police have established similar links with King's College, London. The Forensic Science Society also offers its members Diplomas in Crime Scene Investigation, Fire Investigation, Firearms Examination, and Document Examination. The Forensic Science Society diploma programs are accredited by the University of Strathclyde.

Further links have been established with other universities in the UK and it is anticipated that (distance learning) Bachelor's and Master's degrees in Crime Scene Sciences will be available in the near future. One other link, recently created, is that with Birmingham University. A joint postgraduate certificate course in forensic archaeology is currently being planned.

The UK police service is also considering the adoption of work-based competency qualifications, known as national vocational qualifications (NVQs). A pilot scheme is currently running in six police forces and, if adopted, it is anticipated that crime scene examiners will undertake an NVQ program in their first year of operational service. A parallel program for fingerprint officers is also being proposed.

Contemporary Issues

Another interesting recommendation emanating from one of the 1980s reviews was the unusually high (in comparison with other jurisdictions) 16-point fingerprint standard. The research revealed that the 16-point standard was in fact based on something of a myth. It appears that the original Scotland Yard 12-point standard was raised to 16 points in 1923, following the discovery of a nineteenth century paper written by no less a character than Alphonse

Bertillion of Paris. It appears that Bertillion (not renowned as a fingerprint enthusiast, as he perceived fingerprint technology as a competitor to his system of bertillionage) had masked parts of two dissimilar fingerprints, leaving only similar characteristics available for comparison. Scotland Yard took the decision to err on the side of caution, and thus raised the standard to 16. Having established that the standard had little, if any, scientific foundation, the researchers recommended that the UK should move to a nonnumeric standard, based only on the expertise of the fingerprint examiner and the quality of the mark in question. It was further recommended that quality assurance and quality control systems would support the proposed change. These recommendations were accepted in total and, with effect from April 2000, the United Kingdom will convert to the new, nonnumeric standard.

DNA technology continues to develop month by month and, in common with other agencies throughout the world, the (UK) Forensic Science Service is conducting research into new, low-copy techniques. It is now possible to obtain DNA profiles from objects that have only been touched by an offender. While these are exciting developments with enormous potential for the crime scene investigator, the question of contamination and the consequent management of exhibits and crime scenes is certain to become an ever-increasing problem.

In 1991 the Forensic Science Service moved from being a core Home Office department to 'agency status'. In practical terms this meant that instead of forces receiving (effectively) a free service, they are now required to pay for every examination carried out by the service. However, it also meant that the police market was opened up to other competition and, in the years since 1991, several commercial laboratories have set up in competition to the Forensic Science Service. Forensic science is an expensive commodity and police forensic budgets are, therefore, substantial. Forces have become very skillful in ensuring that they obtain good value for their money. However, the relationship between the Forensic Science Service and the police service has been somewhat strained on occasions and it is fair to say that the move to agency status has been largely unpopular within police circles. Despite these problems, the Forensic Science Service continues to dominate the market. While many practitioners would like to see a return to the pre-1991 relationship, there appears little political will at present to undertake any U turns.

While all fingerprint examiners are trained to undertake manual searches of fingerprint record systems, most bureaux have some form of automated (computerized) system capable of comparing crime scene marks against records and against each other. Having searched the mark, the computer produces a series of 'hits', which are then compared manually by a fingerprint expert. The most common system is known as AFR (automated fingerprint recognition system). Over the past 7–8 years, however, the Home Office has been developing a new, high-quality national system, known as NAFIS – the National Automated Fingerprint Identification System. Bureau staff have now been trained to operate the new system, which is currently being introduced.

Volume crimes, burglary, damage, and auto crime, are issues very close to the heart of the general public. Indeed, most citizens have been affected by incidents of this nature at one time or another in the course of their lives. In an effort to maximize effectiveness and efficiency, several forces are committed to examining volume crime scenes for only the three best evidence types: fingerprints, DNA, and footwear marks. In addition to gaining high-quality evidence suitable for use in a court of law, information gained through these initiatives is also subjected to investigative analysis (intelligence) in an effort to identify trends and individual offenders. Several manual and automated systems have been developed in recent years to manage footwear images (e.g. Shoefit and SICAR), and most forces have adopted some form of footwear comparison systems.

Information technology-related offences, such as pedophilia and Internet fraud, are now very common in all developed countries. This is certainly the case in the United Kingdom. Forces have not been slow to appoint in-house experts capable of recovering and analyzing computer evidence found in the possession of suspects and offenders. National guidelines have been drawn up and published on how computer systems should be seized and managed by crime scene examination personnel. This is clearly a growing problem and the surrounding issues are subjected to constant review.

Following on from the findings of a Royal Commission (government inquiry) into previous miscarriages of justice, the Government has recently initiated the development of a 'Register of Forensic Practitioners'. A Council for the Registration of Forensic Practitioners has now been formed and a chief executive appointed. Professional assessment panels looking at the disciplines of crime scene examination, fingerprint examination, forensic medical examination, forensic pathology, and forensic science are to be formed in the near future. This is a dramatic and exciting development that will clearly raise the profile of the members of the various professions. While the council has not yet set requirements for entry on to

the register, it is assumed that professional qualifications and experience are likely to come into the equation at some stage. Generally speaking, this initiative is viewed favorably by the vast majority of practitioners, who see it as a step in the right direction. It is interesting to speculate on the future status of nonregistered professionals operating in the forensic science field. Will they, for example, be allowed to take part in legal proceedings? Indeed, will they be allowed to practice?

Conclusions

The discipline (or profession?) of crime scene examination has developed dramatically since those heady days when Bertillion and his contemporaries broke new ground over 100 years ago. It would be fair to conclude that while forensic and investigative techniques progressed significantly in the first half of the twentieth century, the advances of the past 20 years or so have been positively breathtaking.

In the past, the primary role of the crime scene investigator has been to assess, record, preserve, and recover. The role is now much bigger than that, with practitioners increasingly being required to assess and interpret evidence. This begs the question as to whether crime scene investigators will, in due course, become experts in their own right. Many observers believe they will. Arguably, the discipline has been transformed from a skill to a science. The demand in every jurisdiction for hard (physical) evidence has never been stronger, and intelligent, highly trained professional crime scene investigators are in great demand.

It should be remembered that in Europe and the United States investigators generally have the luxury of ready access to crime scene examiners and other scene experts. This is not necessarily the case in many other jurisdictions where the examiner is required to have a broad knowledge of the subject and be able to turn his or her hand to the many and varied forensic procedures. The interesting and varied developments outlined in this article are not the sole preserve of the United Kingdom. Similar advances are occurring all over the world. Crime scene examiners (investigators) are now better trained and better qualified than ever before. There is every reason to suppose that this trend will continue. If the discipline is not regarded as a true profession today (alongside that of, say, the forensic scientist), it will almost certainly be so in the future.

See also: **Education**: An International Perspective. **Fingerprints (Dactyloscopy)**: Standards of Proof; Automated Methods, including Criminal Record Administration. **History**: Forensic Medicine.

Further Reading

Browne DG and Tullett BV (1952) *Bernard Spilsbury, His Life and Cases*. London: Odhams Press.

Cuthbert CRM (1958) *Science and the Detection of Crime*. London: Hutchinson.

Eckert WG (ed) (1997) *Introduction to Forensic Sciences*, 2nd edn. Boca Raton, FL: CRC Press.

Evett IW and Williams RL (1996) A review of the sixteen points fingerprint standard in England and Wales. *Journal of Forensic Identification* **46**: 49–73.

Fletcher T (1986) *Memories of Murder*. London: Grafton.

Forensic Science Service (1996) *Using Forensic Science Effectively*. Birmingham: Forensic Science Service.

Forensic Science Service (1999) *Scenes of Crime Handbook*. Birmingham: Forensic Science Service.

Lambourne G (1984) *The Fingerprint Story*. London: Harrap.

Simpson K (1978) *Forty Years of Murder*. London: Harrap.

Thorwald J (1965) *The Marks of Cain*. London: Thames and Hudson.

Thorwald J (1966) *Proof of Poison*. London: Thames and Hudson.

Wambaugh J (1989) *The Blooding*. New York: Bantam.

White P (1998) *Crime Scene to Court: The Essentials of Forensic Science*. Cambridge: Royal Society of Chemistry.

Williams J (1991) *The Modern Sherlock Holmes: An Introduction to Forensic Science Today*. London: Broadside.

Fingerprint Sciences

P Margot, University of Lausanne, Lausanne, Switzerland

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0567

Introduction

Fingerprints and fingermarks combine to provide the most powerful means of personal identification available to police and courts. How this became such a powerful tool arises from the realization that the patterns of ridge skin and their details are unique, immutable, universal, easy to classify, and leave marks on any object handled with bare hands. The

understanding of the value of fingerprints led to research in detection techniques and to operational and strategic uses for fingerprints. This historical overview covers only the forensic science/identification perspective, with a selection of highlights that seemed essential for the development of fingerprint science. It excludes esoteric and artistic uses of fingerprints.

Early History: Empirical Period

There is evidence that early men believed, or even knew, that the complex ridge designs on the surface of the skin of their fingers were an expression of their identity. These beliefs have led to esoteric use with palmistry.

Fingerprints or impressions are found in numerous archeological findings, indicating that they were used as a means of signing artworks (pottery, hand stencils, cave petroglyphs, tiles), constructions (clay brick, e.g. walls from work done under the second Dynasty of Ur some 5000 years ago in Bur Sui) or, later, documents in Japan and China, followed by India. It must be said that in most cases no ridge patterns can be observed.

It is difficult to determine precisely when fingerprints were used for the first time as a means of personal identification in criminal proceedings. A palm print, because of its size, made it possible to exonerate a suspect in Rome in the first century of our era, according to Quintilianus (AD 92). On the other hand, the apostle Paul signed his writings with his fingerprints, as evidenced in the Bible. Even if individuality of the pattern was known or supposed, it can be said that signing by applying an inked finger was more of a symbolic gesture, fraught with belief.

It was not until the seventh century, however, that fingerprints found an intentional documented application for the identification of legal documents in China under the T'ang dynasty (AD 618 – 906), even if this usage seems to go back to the Han dynasty (206 BC to AD 220). Around this time, domestic laws required signatures with a fingerprint in Japan, when parties could not write. But most ancient prints show poor details and could hardly be attributed to a person. Although this may demonstrate the symbolic nature of fingerprints, there is evidence in fictional literature of the time (referring to criminal investigations and court proceedings) that fingerprints were used in identification. There are even mentions of fingermarks from crime scenes: an anecdotal story describes the demonstration of the presence of the favorite concubine of emperor Ming Huang (713–756) in a given location of the palace through her fingermark.

The Scientific Early Days

Interest reappeared in modern times after the Renaissance. Nehemiah Grew, an English botanist, described details of the sudoriferous pores, the papillary ridges and their patterns in 1684, and is claimed to be the first scientist to clearly document his findings. Grew never referred to the uniqueness of friction skin patterns. With the development of modern medicine and anatomy, the Italian anatomist Malpighi made a passing mention of the patterns on ridged skin (1686) while studying the function of the skin as a tactile organ and its use in the enhancement of traction for walking and grasping. A deep layer of the skin still bears the name 'malphigian layer'.

Johan-Evangelist Purkinje (1787–1869)

In the eighteenth century and early nineteenth century, several works of anatomy were based on the study of dermatoglyphs, and Purkinje, a Bohemian-born Czech, published his doctoral thesis or *Commentatio* entitled: *De examine physiologico organi visus et systematis cutanei* (On the physiological examination of the organs of vision and the cutaneous system) in 1823, the year when he became professor at the University of Breslau. This is the first published work that describes and classifies fingerprint patterns into nine groups. Interestingly enough, Purkinje never saw the potential applications of these patterns and did not publish anything else on the subject. It is not surprising that this work remained obscure for almost 60 years.

In 1880, Faulds proposed the use of fingerprints for the 'identification of criminals from the traces left at the crime scene' in the journal *Nature*. Herschel claims priority for this proposition; however, his proposal only involved fingerprints as a means of 'identification of habitual offenders'.

William James Hershel (1831–1907): the identification of habitual offenders

Hershell started to have an interest in fingerprints as a young civil servant in Bengal in 1858, possibly after observing local rituals and finding the strength of engagements binding Indians who signed documents using handprints. Throughout his career he took collections of fingerprints and studied them as a hobby until his appointment as a magistrate and tax collector in Hooghly in Bengal, near Calcutta. Through his practice and with time, Hershel noticed what had, until then, escaped scientists and anatomists: finger ridge patterns were an identifying sign *par excellence*. The prints of a person remain identical through time and differ from person to person.

This led him to introduce a system of keeping records in legal registers, and finally he introduced fingerprints into the prisons to avoid substitutions. Convinced that the patterns were unique and individual in their details, as well as immutable, he wrote a letter in England to the Inspector of Jails and the Registrar-General to offer what he described as an effective and efficient new system for the identification of recidivists. This letter, which is often referred to as 'the Hooghly letter', of 15 August 1877, did not generate any response at the time and was filed away.

Dr Henry Faulds (1843–1930): identification of criminals by fingermarks

Faulds, a medical missionary from Beith in Scotland, took an interest in fingerprints while in a mission in Japan. As a hobby, he studied prehistoric pottery and observed fingerprints on many items. This led him to take an interest in fingerprint patterns and to compare fingerprints from humans and monkeys. There might have been a desire to find evidence contradicting Darwin's evolution theory because Faulds initially wrote to Darwin in February 1880, before publishing a letter in *Nature* entitled 'On the Skin-Furrows of the Hand' on 28 October 1880. This letter is fundamental. It indicated for the first time the possibility of identifying criminals by papillary marks left at crime scenes. It went even further, in indicating two cases where a practical dactyloscopic demonstration was used: one in identifying the author of a mark on a glass; the other excluding a suspect (the exclusion was based on sooty marks on a white wall). Faulds also described how to take comparison prints and how to compare prints by using juxtaposed projections. This indeed encapsulated almost everything dactyloscopy was to offer in the twentieth century!

Francis Galton (1822–1911): scientific bases for fingerprint use

Darwin, in poor health, forwarded Faulds' letter to his cousin Francis Galton, who filed it and apparently thought no more of it until he was asked to give a lecture on the anthropometric system of Bertillon (or bertillonnage), as an evening lecture at the Royal Institution in London in 1888. An independent and scientific mind, Galton reflected on other methods of identification and remembered reading and corresponding about fingerprints. He contacted Herschel, who helped him with much material collected over the years. Galton then realized the importance of this development and took a deep interest in the matter. This was the beginning of the true formalization of dactyloscopy. The scientific studies led by Galton allowed criminalistic applications for fingerprints. The results of his studies include embryology, morphology, genetic inheritance, permanence, individuality, as well as an historical background, methods of printing and potential usefulness. This was published in 1892 under the title *Finger Prints* and is considered a landmark treatise. It can be associated with the beginning of the modern use of ridged skin prints and marks as a means of identification.

One of the problems that remained to be formalized was the setting up of an encoding and classification system that allowed the searching of files containing thousands and more records, tasks that were perfectly fullfilled using anthropometry.

Juan Vucetich (1858–1925) and Edward Richard Henry (1850–1931): classification

A classification system appeared in Argentina in 1891, designed by Juan Vucetich, a native of Dalmatia, who produced a clear and simple filing system that was in use in many Latin countries until recently. Vucetich recognized that Galton was the driving force and inspiration behind his system. He coined the name *icnofalangometria* for his system, but it is a visitor to Vucetich, Dr Francisco Latzina, who proposed replacing this inaccurate name by *dactiloscopia* in a report in the journal *La Nacion* in Buenos-Aires.

It is worthwhile noting that the so-called Rojas Murder (1892) was solved when a fingermark in blood on a door was associated with the right thumb of the author Francisco Rojas, by inspector Alvarez of the Argentinian Central Police. It comes as no surprise that Argentina was thus the first country to introduce dactyloscopy as a unique means of identification as early as 1896. Other South American countries followed very quickly.

Soon afterwards, an initial combined classification with anthropometry, adopted in London's Scotland Yard, was replaced in 1900 by a system designed by Sir Edward Richard Henry in India. Henry had been in contact with Hershel when still in India and was aware of Hershel's work. He proposed the official introduction of dactyloscopy in India from 1897. The classification was developed with two police officers, Azizul Haque and Chandra Bose. Introduced in Britain, this became the standard for large files and was adopted widely. This system of fingerprint classification remains the most widely used system in the world and, at the end of the twentieth century, had only started to be replaced by computerized classification systems. It introduces a classification based on fixed points (deltas and core) and four basic patterns (arches, loops, whorls and composites). Ridge counting and ridge tracing help reduce these general classes and are obtained by counting the number of ridges crossing an imaginary line drawn between core and delta (so-called Galton line) and looking at the relative

position of ridges originating from one delta against a second delta.

Numerous other systems had local applications and were never adopted widely, but all these systems were sufficiently incompatible for the major issue of the beginning of the century to be that of standardization.

Early systems were based on a classification using general patterns from all 10 fingers; they were the so-called 'decadactylar' (or 10-print) systems. Much later on, classifications based on single fingers were designed for very large files: 'monodactylar' (or single-print) systems (Oloriz in Spain, Borgerhoff or Stockis in Belgium, 1914). Some classifications for palms were also made: 'palmar' classification.

For many years, Bertillon in Paris championed his anthropometric system against dactyloscopy and rejected the latter for lack of sufficient diversity for an identification system. One can find an addendum of one page about dactyloscopy in his book on criminal anthropometry published in 1893 (second edition). It is interesting to know that while Bertillon opposed the introduction of dactyloscopy in France, a very active group around Lacassagne in Lyon became what is sometimes referred to the 'Lyon school', from which came later the famous French criminalist Edmond Locard. Major steps were made there to detect latent prints and use them for identification purposes. The names of Florence and Coutagne were well known in forensic medicine circles; Frécon and Forgeot wrote two theses on detection and exploitation of marks in 1889 and 1891. A Prussian official, Eber, proposed the introduction of systematic detection of fingermarks and developed what could be considered as intervention kits for latent mark detection in 1888.

Despite Bertillon's resistance, dactyloscopy became the main means of identification for police work in the replacement of anthropometry at the turn of the century, but in Paris it was not until Bertillon's death in 1914 that dactyloscopy was finally adopted as the single most powerful means of identification. This coincided with the First International Meeting of Police in Monaco. This was a key meeting for standardization, although most decisions were postponed owing to World War I; this was the first seed of what was to become Interpol.

Dactyloscopy constitutes to this day the evidential proof of identity par excellence. It is rather ironic that one of the first identifications obtained with fingermarks as sole evidence was by Bertillon himself in 1902 in the case against Scheffer!

Later Developments

Several noteworthy events should be mentioned in relation to the progress of dactyloscopy, and it is difficult and a question of choice as to whether or not some events can be considered as landmarks. The selection is therefore somewhat arbitrary. It concerns three main areas: one fundamental is the discussion of the probative value of fingerprints in the identification process; the second concerns operational needs in the detection of latent fingermarks on various types of surfaces and under various environmental conditions; and the third concerns operational and strategic needs in the way fingerprints are being used (databases, searching, etc.).

Probative value of fingerprints in the identification process

Fingerprint patterns are unique and individual for a given finger because the detail in a fingerprint is formed in an accidental manner during gestation. (Although two fingerprints may be similar, it was demonstrated by Katherine Bonnevie that even identical twins had different fingerprints.) Inez Whipple (1904), Wilder and Wentworth (1918), followed later by Cummins and Mildo (1943) and Alfred Hale (1952), were pioneers in the science of dactyloscopy and gave numerous observations on the formation of papillary ridges. From the third month of fetal life, when volar pads are regressing, bumps or points (ridge units) develop at variable rates on the dermal surfaces, with each unit containing a pore. These fuse together as lines to give the friction ridge. The general pattern is influenced by heredity but this is not the case for papillary minutiae, as these are the result of stress and variable pressure on the volar pads when the units fuse into lines. The ridge characteristics that make fingerprints unique are known as minutiae, Galton points, characteristic points or points of identification. It was also noticed by Locard (1912) that pore distribution, shapes and sizes were unique. All this led to the discussion of what criteria were necessary for a positive identification or individualization. Founded on the basic statistical works of Galton, then Balthazard (1911) and others, the question of the need for a minimal number of points (minutiae) necessary for identification quickly turned, in most countries, into an operational decision on a fixed number (numerical standard). This fixed number varied in different countries but a majority set this limit at 12, based on restricted interpretation of publications by Locard. A few countries considered the numerical standard as the sole deciding factor, whereas other countries considered qualitative factors (clarity of ridges, presence of pores, etc.) and used a lower numerical standard. Locard even proposed the use of partial information for qualified opinions, rejected by a majority of dactyloscopists to this day. Some countries decided on a higher numerical standard, partially based on the

misunderstanding of work by Balthazard (in Italy) or demonstrations by Bertillon. It was in 1953, at a general assembly meeting of Interpol in Oslo, that a report by Santamaria Beltran showed that the absence of minutiae or rare minutiae had identification value that was largely ignored. Fixed numerical standards were seen to be inappropriate for dealing with such data. Following a 3 year study, the International Association for Identification (IAI) resolved in 1973 that 'no valid basis exists for requiring that a pre-determined minimum number of friction ridge characteristics must be present in two impressions in order to establish positive identification.

Although this resolution was adopted at the time by the United States and Canada, it is only since 1997 that other countries have begun to adopt this position, following demonstrations of the statistical variability of patterns and details and the subjectivity of decisions taken by fingerprint experts in blind testing. This is leading dactyloscopy to a new era of quality management schemes.

Detection of latent fingermarks

At the very beginning of modern dactyloscopy, early authors proposed numerous methods for the detection of latent marks – powders, iodine fumes, silver nitrate, etc. – and many empirical developments followed with mixed success. Early analyses of secretions from the skin did not lead to spectacular advances until 1954, when ninhydrin, a reagent for amines, very sensitive to amino acids, was introduced for the detection of fingerprints on paper by two biochemists Odén and von Hofsten. This was an important step because porous surfaces like paper were always difficult surfaces to examine, with little chance of detecting a quality print more than a couple of days old. Then in the late 1970s the observation and detection of photoluminescent fingerprints by Dalrymple and coworkers in 1977, using lasers, allowed a quantum leap in the sensitivity of fingermark detection. Observation in 1978, by the Metropolitan Police of Tokyo, that cyanoacrylate glues, when vaporized, polymerize preferentially on fingerprint deposits, led to a widespread laboratory technique to fix and detect marks on smooth surfaces, with many possibilities for staining and enhancing contrast. The introduction, by the Home Office in Great Britain, of detection techniques, such as physical developer, vacuum metal deposition, radioactive labelling and, more recently, 1,8-diaza-9-fluorenone (DFO), led to extensive research by many groups internationally to find ways of increasing sensitivity and selectivity of detection of latent marks on various surfaces and/or exposed to various environmental conditions. This surge in research is still going on today.

Operational and strategic uses for fingerprints

The classification of fingerprints into distinct groups based on general similarities allows the fingerprint examiner to search for an unidentified fingerprint/mark within a specific section of the fingerprint file, rather than having to search the whole file.

Computer systems that entirely automate the storage and searching of fingerprints (AFIS: Automatic Fingerprint Identification Systems) have seen the decline of traditional classification methods at national levels. Three principal AFIS systems are currently in use in a number of countries after early developments carried on in the United States under the aegis of the FBI in the late 1960s. Introduction on a large scale of such systems started in the 1980s and they have demonstrated their power in identifying authors of crimes. Although it was a theoretical possibility to have 'cold hits' (i.e. identification of persons in cases without any suspect) with traditional systems, it is only with the modern AFIS that regular cold hits became the real operational gains that were hoped for even in the early days of fingerprinting.

See also: **Crime-scene Investigation and Examination**: Fingerprints. **Evidence**: Statistical Interpretation of Evidence/Bayesian Analysis. **Fingerprints (Dactyloscopy)**: Visualization. **Identification/Individualization**: Overview and Meaning of ID.

Further Reading

Asbaugh DR (1996) *Quantitative-Qualitative Friction Ridge Analysis*. Hope, BC: Ridgeology Consulting Services.

Berry J (1991) The history and development of fingerprinting. In: Lee HC and Gaenslen RE (eds) *Advances in Fingerprint Technology*, New York: Elsevier.

Champod C (1996) *Reconnaissance Automatique et Analyse Statistique des Minuties sur les Empreintes Digitales*. PhD thesis, University of Lausanne.

Champod C and Margot P (1996) *L'Identification Dactyloscopique*. Lausanne: University of Lausanne.

Cowger JF (1983) *Friction Ridge Skin*. New York: Elsevier.

Galton F (1892) *Finger Prints*. New York: Da Capo Press, (Reprint edition 1965).

Heindl R (1927) *System und Praxis der Daktyloskopie und der sonstigen technischen Methoden der Kriminalpolizei*, 3rd edn. Berlin: De Gruyter.

Lambourne G (1984) *The Fingerprint Story*. London: Harrap.

Locard E (1931) Traité de Criminalistique. Vol. I: *Les Empreintes et les Traces dans l'Enquête Criminelle*. Lyon: Desvignes.

Margot P and Lennard C (1994) *Fingerprint Detection Techniques*, 6th edn. Lausanne: University of Lausanne.

Forensic Anthropology in the USA

K A R Kennedy, Ecology and Evolutionary Biology, Cornell University, Ithaca, NY, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0786

Overview

In 1972 forensic anthropology came of age in the United States when the American Academy of Forensic Sciences (AAFS, established in 1948) recognized this subdiscipline of physical anthropology as a separate section comprising members qualified to serve as expert court witnesses in medicolegal cases by virtue of their experience in research and education. The Forensic Sciences Foundation (FSF) became the research and training branch of the AAFS in 1973. This specialized and applied aspect of human skeletal biology and osteology, known as 'forensic anthropology' as early as the 1940s, established the American Board of Forensic Anthropology (ABFA) five years after the founding of the Section with its responsibility of certifying qualified MD and PhD candidates as Diplomates following their successful performance in practical and written examinations administered by Fellows of the Section. Today the Physical Anthropology (Forensic Anthropology) Section represents one of the major forensic sciences; the other Sections are designated as Criminalistics, Engineering Science, Odontology, Pathology and Biology, Psychiatry and Behavioral Sciences, Toxicology and a General Section. There are annual meetings of the AAFS and its organ of publication is the *Journal of Forensic Sciences*.

Personal identification of human bones, teeth, hair, nails and soft tissues (*partis mollis*) within the medicolegal sphere has a history in the United States beginning in the late colonial-revolutionary period. Based on milestones in research and publication documentation, historic periods may be categorized as (1) late eighteenth century to the last quarter of the nineteenth century; (2) 1878–1939; (3) post-World War II to the last quarter of the twentieth century; (4) 1972 to the present. These time frames are recognized in the historical studies of forensic anthropology by various American scholars.

Late Eighteenth Century to the Last Quarter of the Nineteenth Century

An early case of personal identification of human remains was that associated with Paul Revere (1735–1818) who, when not engaged in midnight rides, was occupied with silversmithing and fashioning dentures from hippopotamus bone and sheep teeth. In the early 1770s he made two teeth for the mouth of the Boston physician and military officer Joseph Warren (1741–1775). Warren spied against the British and met his demise at the Battle of Bunker Hill in 1775. The victorious British placed his body in an unmarked grave (not out of disrespect but to preserve the remains from animal predators scavanging the battlefield). A year later Warren's brother John (1753–1815) disinterred the body, but how could he be certain it was Joseph? It is said Revere who had been brought into the identification case was able to identify Joseph as he recognized his own handiwork in the two teeth in the skull.

A more significant event in early forensic anthropology involves the court testimonies of the Harvard University physicians and anatomists Oliver Wendell Homes (1809–1894) and Jeffries Wyman (1814–1874) in 1850 at the George Parkman murder trial in Boston. The victim, an alumnus and benefactor of the university, was slain in the laboratory of chemistry Professor John White Webster (d. 1850) who disposed of segments of the dismembered body in his office privy, a tea chest and a stove. A vigilant janitor of the building unlocked Webster's lab, discovered bits and pieces of human flesh in unlikely places, even for an area adjacent to a dissection room, then informed the police whose search revealed additional body parts including a burnt skull and damaged dentures. These were examined by Homes, Jeffries and other members of Harvard's medical staff. Before his appointment with the hangman, the chemistry professor confessed to the crime which was motivated by his failure to repay Parkman some long outstanding monetary debts.

Not only was Webster the first Harvard Professor to be hung for his sins, but the case is also unique in its involvement of 'expert witnesses' setting before the court the results of their thorough examination of human remains in order to determine if these are human or nonhuman, the number of individuals present, determination of age and sex, time elapsed since death and dismemberment, presence or absence of chemical substances and manner of death – all components of the protocol of a modern forensic anthropology analysis. This notorious case alerted the public and law enforcement–jurisprudence professionals that human remains could be identified, but the discipline of forensic anthropology suffered from retarded development for several more decades. However, innovations in other areas of anthropological investigation emerged in the nineteenth century and contributed to forensic anthropology. Among these was the enthusiasm over worldwide human racial identification and classification, the principal

research goal of ethnographers and physical anthropologists during this historic period. Numerous anthropometric instruments were invented for gaining mensural data on both skeletal and living subjects.

The discovery of a hominid fossil record in Europe beginning in 1829 led to the classification of prehistoric human ancestors as anatomically modern (Cro-Magnon) or anatomically archaic (Neanderthal) *Homo sapiens* and to the establishment of the field of paleoanthropology (human paleontology and prehistoric archaeology). These research orientations fostered the growth of human osteological collections, a pursuit already advanced by phrenologists whose popular societies in Europe and the United States focused on discernment of behavior, race and social ranking from observations of cranial morphological and mensural variations.

Phrenology has joined the dustbin of criminal anthropology and other pseudosciences, the latter based on the theory of the Italian criminologist Cesere Lombroso (1836–1909) that particular physical features of the living human body or skeleton were atavistic. These hereditarily degenerative stigmata distinguished criminal types from law abiding folk. Lombroso compared social deviants with Neanderthals in his use of anthropometric and morphological observations, his criteria offering the prospect that evildoers could be removed from the streets before they had committed their crimes! Some scholars claim that criminal anthropology had minimal influence in the development of modern forensic anthropology, but this topic was taken seriously by the Harvard anthropologist Earnest Hooton (1887–1954) as late as the 1930s, and he trained a number of students who became forensic anthropologists.

Criminal anthropology should be distinguished from the methods of the French anthropologist Alphonse Bertillon (1853–1914) who selected anthropometric measurements for the purpose of human identification since values for a broad spectrum of measurements are admittedly unique to each individual (although alterable because of changes in age and disease). By 1920 the Bertillonage System was replaced in police departments and penal institutions by fingerprinting, but this earlier method of personal identification was based on traditional anthropometric methods used in racial anthropology and human paleontology and is a part of the history of modern forensic methods used by anthropologists and molecular biologists developing DNA identification procedures.

The Next Sixty Years: 1878–1939

By the late nineteenth century physical anthropology had gained respectability within the international scientific community hence it is not surprising that research on individuation and estimations of age, sex, race and pathological markers of the skeleton meant that a number of anthropologists contributed their expertise to medicolegal cases. For example, stature reconstruction formulae were devised by the 1880s, and racial anthropologists were listing physical sorting criteria in their typological efforts at classification of humanity, prehistoric and modern. The founding of the Société de Anthropoologie de Paris in 1859 by Paul Broca (1824–1880) was one of the earliest learned societies with members contributing research in the publications of scientific journals they sponsored. In the United States the Anthropological Society of Washington was founded in 1879 out of which developed the American Anthropological Association with its first issue of the *American Anthropologist* appearing in 1888. Section H (Anthropology) of the American Association for the Advancement of Science (founded in 1851) was formed in 1882. Publication of the *American Journal of Physical Anthropology* appeared in 1918 and the American Association of Physical Anthropologists held its inaugural meeting in Charlottesville, Virginia, in 1930. These and other organizations and publications advanced research in areas of knowledge and practice known today as 'forensic anthropology', but if a 'father' of the field in the United States can be identified it would be Thomas Dwight (1843–1911). As Parkman Professor of Anatomy at Harvard (and Holmes' successor to that chair) and with training in anthropology, Dwight (1894) published and taught subjects relevant to personal identification and human skeletal biology, e.g. stature estimations, racial variation and aging modifications. In 1878 he won a prize from the Massachusetts Medical Society for his essay 'The Identification of the Human Skeleton: A Medical-Legal Study', a landmark in the emergence of forensic anthropology out of its traditional home in medicine and anatomy.

During his forty years of teaching, Dwight had communication with George Amos Dorsey (1869–1931) who received a PhD in anthropology at Harvard. As a consequence of reading Dwight's (1894) Shattuck Lecture paper, Dorsey applied some of the methods of human identification to Native American skeletons housed in the Field Columbian Museum at Chicago where he became curator. His contributions to forensic anthropology might have continued into his later years but for the negative evaluation of his court testimony by opposing medical colleagues also connected with the Luetgert trial in 1897. Adolf Luetgert was a sausage manufacturer in Chicago who, in a moment of pique, decided to dispose of Mrs Luetgert in one of the meat-processing vats at his establishment. The rumor that his victim became an

ingredient in his delicious bratwurst remains unsubstantiated, but discovery by police of bone fragments on the premises did not enhance the advancement of Mr Luetgert's business. His sausage-making came to an end when Dorsey identified the bone fragments as human. Despite the success of his lecture to the Medical–Legal Society of Chicago following the trial Dorsey (1899) abandoned forensic anthropology and went on active duty with the US Navy in World War I.

Despite the public furor over Luetgert's conviction, physical anthropologists of the pre-World War II period did not respond immediately by creating forensic anthropology as one of their subfields of study. Hooton's courses on human evolution and diversity at Harvard excluded forensic training beyond general osteology. An appendix in his popular book *Up From the Ape* includes some discussion of methods for age, sex and race determination. Nor did this academic publish reports of the occasional medical–legal cases that came his way. Hooton's contemporary, Alex Hrdlicka (1869–1943), was a Bohemian-born anthropologist who became Curator of the US National Museum (Smithsonian Institution) in 1910. He too was consulted by law enforcement personnel. His cases are unpublished. However, these two prominent anthropologists trained a generation of students who came to identify themselves as forensic anthropologists. Hooton's students include Alice Brues, J. Lawrence Angel (1915–1986), Harry L. Shapiro (1902–1990) and Frederick S. Hulse (1906–1990); Hrdlicka's apprentice and successor at the Smithsonian was T. Dale Stewart (1901–1997).

Another academic of this period who wrote about forensic anthropology and was a teacher was Harris Hawthorne Wilder (1864–1928). an American who received his doctoral degree in zoology from the University of Freiburg, Germany, and later held a professorship at Smith College in Massachusetts. Wilder's book, co-authored by a police expert in identification, was entitled *Personal Identification: Methods for the Identification of Individuals, Living and Dead* and it was the major source on the subject before the outbreak of World War II. A peripheral figure who was a contemporary of Wilder's was the American anatomist Paul Stevenson (1890–1971) whose career in China led to publications about age and racial variations of long bone development among Far Eastern populations, but it is uncertain if the author understood the forensic implications of his work.

More significant advances in forensic anthropology took place at Case Western Reserve University in Cleveland, Ohio, where the anatomist T. Wingate Todd (1885–1938) curated and expanded skeletal collections first organized by Carl August Hamann (1868–1930), Professor of Anatomy at Case since 1893 and Dean of the Medical School in 1912. Skeletons were salvaged from dissection rooms and every cadaver was accompanied by a record of its age at time of death, sex, race, place of birth, occupation, medical history and cause of death. The Hamann-Todd collection contains over 3000 individuals. Using this series, Todd published extensively on determinations of age from changes in the pubic symphysis and cranial suture development. Todd trained a number of major contributors to forensic anthropology and included the subject in his lectures. Wilton M. Krogman (1903–1988) worked in Todd's laboratory and on the basis of his studies of the skeletons in the collection he was able to reconstruct the morphometric profiles of two skeletons from a Native American burial deposit, his publication of results being widely read by his peers.

A second well-documented osteological collection was established by Robert J. Terry (1871–1966) at the Department of Anatomy of Washington University in St. Louis, Missouri. The collection is housed today at the Smithsonian Institution. Cadavers dissected from the period 1914 to 1965 bring the size of the series to 1636 individuals each one with excellent documentation. These were studied by Mildred Trotter (1899–1991) of Washington University and provided the data base for her later research on reconstruction of stature of American war dead of World War II and the Korean conflict.

Post-World War II to the Last Quarter of the Twentieth Century

Prior to World War II forensic anthropology was a peripheral activity of those physical anthropologists willing to assist in law enforcement investigations. The majority held academic positions and they contended with the bias of many of their colleagues that 'police work' and newspaper publicity were incompatible with life in the ivory tower. Not surprisingly, reports of their analyses of decomposed bodies and skeletons were not published in scientific journals, not even in the *American Journal of Physical Anthropology* prior to 1940, although articles on osteology abound. Nor did these academic forensic anthropologists receive financial compensation for their contributions to medical–legal inquiries. There were no training programs and field schools in forensic anthropology in the United States at this time.

Two approaches to the practice and teaching of forensic anthropology are discernible in this period. First, teaching of human osteology was offered by physical anthropologists at colleges and universities who undertook some forensic work outside of their major research and instructional programs but did

not publish case studies nor receive payment. However, some of these academics produced students whose careers combined laboratory research and consultation with medical–legal clients. Secondly, other physical anthropologists were associated with, or employed by, research and military institutions and were active in publishing their investigations and appearing in court as expert witnesses.

Within the first group – the academics – are included Dwight and Hooton of Harvard University, Wilder of Smith College, Todd of Case Western Reserve University, Hulse of the University of Arizona, James E. Anderson (1926–1995) of the University of Toronto, Georg K. Neumann (1908–1971) of the University of Indiana, Daris Swindler of the University of Washington, and Theodore D. McCown (1908–1969) of the University of California at Berkeley. Only a few of these scientists ever appeared in court as expert witnesses, and it is interesting that McCown's cases all involved identification of the skeletal remains of the 'eminent dead' (Father Junipero Serra, Juan Bautista de Anza, Amelia Earhart). Krogman (1903–1987), who held positions at Case Western Reserve University, University of Chicago, and University of Pennsylvania at different times in his life, was a forensic anthropologist who may be placed in both categories, as may perhaps Todd, since they held research positions in medical facilities while teaching as members of university faculties. Brues and Trotter held academic positions after their work with the US military.

The awareness of the US Armed Forces that physical anthropologists could use their skills for identification of service dead of World War II and the Korean Conflict brought forensic anthropology into sharper focus among medical and legal professions as well as among historians. McCown was in military service at the San Francisco Presideo from 1942 to 1945 soon after receiving his PhD at Berkeley. His duties included identification of skeletal remains of American war dead from the Pacific theatre, investigations he continued from 1948 to 1950 while he was on the faculty of the Department of Anthropology of his *alma mater*. In collaboration with officers of the Memorial Division of the Office of the Quartermaster General and with a former student, Russell W. Newman (1919–1981) who was head of the Anthropology Unit, McCown directed a color–sound film on the subject of methods of exhumation of human remains. This was screened for military recovery teams charged with the location and removal of skeletalized war dead for repatriation and reburial. Also engaged in skeletal identification in the Pacific war zone was Charles E. Snow (1910–1967) of the University of Kentucky who was invited by Francis E. Randall (1914–1949) of the Anthropology Unit, Research and Development Branch of the Quartermaster General to assist in the Central Identification Laboratory in Hawaii (SILHI) after its establishment in 1947. Snow was replaced by Trotter who refined the earlier stature reconstruction procedures of Rollet to include American military service personnel of European, African and Puerto Rican ancestry. During World War II the American Graves Registration Service was assisted by European colleagues through a branch of the Central Identification Point at Strasbourg where American war dead were brought. Shapiro at the American Museum of Natural History in New York was an advisor to these teams.

The Korean Conflict of 1950–1953 brought together at Kakura, Japan, forensic anthropologists who were major scholars from academic and research institutions: Stewart, Kerley, Paul Baker, Newman, Charles P. Warren. Both publications and seminars resulted from this second endeavor to use the expertise of physical anthropologists in personal identification within the military sphere, studies of age changes of the publc symphysis being especially innovative. During the Vietnam War of the 1960s and early 1970s the services of American forensic anthropologists were carried out at Saigon and in Thailand. Present teams are at SILHI in Hawaii, Wright-Patterson Air Force Base in Ohio, and the US Army Natick Research, Development and Engineering Center in Massachusetts.

This historic period is marked by the Federal Bureau of Investigation's (FBI) interest in the applications of physical anthropology to the military setting as well as Krogman's (1939) publication of *Guide to the Identification of Human Skeletal Material* published in 1939 in the *FBI Law Enforcement Bulletin*. This synthesis of forensic anthropological research was enhanced by Krogman's own studies of age changes of the skeleton, knowledge he acquired under Todd's direction at Case Western Reserve University. In 1962 Krogman wrote the first textbook on forensic anthropology which was followed by a second and revised edition in 1986. These publications had profound influence in shaping and directing the field of forensic anthropology. Furthermore, Krogman's contribution to FBI investigations were advanced by Stewart at the Smithsonian Institution. Stewart had worked with Hrdlicka on various unpublished cases and in 1942 when he was appointed curator his case load with the FBI intensified. Stewart advanced to Director of the Smithsonian in 1962, the FBI connection being continued by Angel who published his case studies and new research methods. Krogman, Stewart and Angel built the bridge between the anthropological and medical–legal communities by demonstrating that the skills of the physical anthropologist engaged in

skeletal biology and paleodemography could be applied to needs of society by personal identification in a scholarly manner that did not tarnish the walls of the ivory tower. Furthermore, these three scientists challenged other forensic anthropologists to design research projects to enhance success in case studies.

This merger of academic, military and government interest in forensic anthropology is demonstrated by the organization of scientific meetings in which experts in these fields could come together for discussion and debate. The first US symposium on forensic anthropology took place in 1948 at the annual meeting of the American Association of Physical Anthropologists held in Washington, DC. There were only four speakers: Krogman, Stewart, Shapiro and C.E. Snow. Eight years later the Wenner-Gren Foundation for Anthropological Research sponsored a summer seminar on forensic anthropology in New York. By the late 1960s Angel directed a course for law enforcement officers, medical examiners and anthropologists held at the Smithsonian Institution, a program continued by Douglas H. Ubelaker following Angel's death in 1986.

Graduate training in forensic anthropology began at a few universities, perhaps first at the University of California at Berkeley with McCown, followed by other training programs at the University of Colorado with Brues and at the University of Kansas with Thomas McKern (1920–1974), Kerley and William Bass.

Multivariate statistical analysis emerged in the late 1950s as an important tool in the protocol of forensic anthropological investigation, particularly for estimation of age, sex, race, and stature. Indeed, by the final quarter of the twentieth century few learned journals or podium and poster presentations at meetings of physical anthropologists failed to have contributions on forensic anthropology without graphs and tables documenting statistical analyses.

1972 to the Present

Fourteen members of the AAFS were present in Atlanta in 1972 when the Physical Anthropology Section was established. Membership in the Section stood at 236 in 1998 of whom 59 (as of 1999) are Diplomates certified by the ABFA. Since 1986, demonstration and maintenance of high professional standards for those passing the rigorous practical and written board certification examination for Diplomate status (DABFA) involves a yearly recertification report of teaching, research, court cases and other activities. Diplomates were involved in 1822 cases in 1997 with 26 court dispositions and 24 testimonies. Over the past decade regional groups of forensic anthropologists have been established: the Mountain, Desert and Coastal Forensic Anthropologists situated in the western United States; the Mountain, Swamp and Beach group in the southwestern states; the North East Forensic Anthropology Association in the northeast and Canada; and the Midwest BioArcheology and Forensic Anthropology group. These organizations provide opportunities for forensic anthropologists from various academic and government institutions to discuss research innovations and case studies.

An increasing number of forensic anthropologists are participating in personal identification at sites of mass disasters (the Oklahoma City bombing, fire at the Branch Davidian compound, airplane disasters of Pan Am flight 103 and TWA flight 843, the Great Thompson flood and military operations of Operation Desert Storm) and recovery of victims of genocide in Bosnia, Kosovo, Argentina and other countries. Many of them work with the National Disaster Medical System and Disaster Mortuary Teams (D-MORT) which have been active since 1993 in response to mass disasters, and with branches of the FBI.

Education in forensic anthropology is offered at a few American universities of which those listed as most prominent in 1982 included the Universities of Tennessee, Arizona, Nevada at Las Vegas, Maryland, Florida, New Mexico and California State University at Fullerton. Some of these have maintained their programs whereas others not mentioned by Snow in 1982 have arisen more recently: Arizona State, Florida Atlantic, Illinois, Cornell, Kansas State, Indianapolis, Michigan State, Western Michigan, Wyoming, San Diego State, Southern Florida, South Carolina, Indiana, State University of California at Chico and Maine. This is not an exhaustive list and changes reflect the retirement, death or transience of forensic anthropology faculty member of whom the majority are in anthropology departments. Some institutions award the PhD degree in anthropology or biological sciences with a specialization in forensic anthropology, whereas others offer only the MA or MSc degree in this subject.

Beyond the ivory tower are courses and programs offered by the US Armed Forces Institute of Pathology in Washington, DC which initiated its training classes in 1988. Four years later an international course sponsored by the Smithsonian was given at Brest, France. The University of Indianapolis has been very active in organizing short field and laboratory courses for university students and professionals in law enforcement agencies and medical pathology. Apart from popular interest in the forensic sciences evoked by the television series *Quincy* and *X-Files*, forensic anthropologists have become heroic figures

in novels by Aaron Elkins and Kathleen Reichs who have both had training in this field. There are edited volumes of case reports and methodologies as well as introductions to forensic anthropology in the form of biographies and individual case studies.

Problems for the future affect younger members of the discipline who must find funding to build teaching and research facilities with osteological collections, often in traditional departments of anthropology in universities where their socioculturally oriented colleagues have not been in a laboratory since they took high school chemistry and may dispute the rationale for surrendering space in an already crowded building. Since the Vietnam war years when 'applied anthropology' became linked to protests against the 'industrial-military establishment' and real or rumored CIA operations, many anthropologists question the political correctness or scholarly merits of any practical applications of their discipline, a question that may ellicit a blunt answer when the forensic anthropologist down the hall is examining a decomposing body or grizzly skeleton. There is anxiety about academic research institutions making sound decisions in the replacement of positions by retiring or deceased forensic anthropologists, although it is possible to lose sight of the fact that new faces appear over the horizon. There are few jobs for forensic anthropologists, except in government and military agencies, hence young PhD students must enter the portals of academe as physical anthropologists, although a significant research and applied interest in forensic anthropology may be found acceptable. It remains the charge of practicing established professionals to get the word to medical examiners and law enforcement personnel that no investigation of human skeletal remains is complete without the contribution of a well-trained and experienced forensic anthropologist. Reliable estimates of age at time of death, sex, stature, pathology, trauma, markers of occupational stress, manner of death, time elapsed since death and features of individuation do not emerge from reading a textbook but depend on years of research experience and varied cases. Current developments in DNA analysis, taphonomy, aging techniques and bone microscopy are welcome additions to the armamentarium of methodology, supplementing the mastery of a knowledge of human anatomy and skeletal biology. Trends in these directions are progressing as forensic anthropology in the United States enters this century as a respected component of the other forensic sciences.

See also: **Anthropology**: Overview; Skeletal Analysis. **Education**: An International Perspective, **History**: Forensic Sciences.

Further Reading

Bass WM (1979) Development in the identification of human skeletal material. *American Journal of Phytsical Anthropology* **51**:555–562.

Dorsey GA (1899) The skeleton in medicolegal anatomy. *Chicago Medical Recorder* **16**:172–179.

Dwight T (1894) The range and significance of variations in the human skeleton: the Shattuck Lecture. *Boston Medicine and Surgery Journal* **13**(1-4):73–76, 97–101.

Hooton EA (1946) *Up From the Ape*. New York: Macmillan.

İşcan MY (1988) Rise of forensic anthropology. *Yearbook of Physical Anthropology* **31**:203–229.

İşcan MY and Kennedy KAR (eds) (1989) *Reconstruction of Life from the Skeleton*. New York: Alan R. Liss.

Kerley ER (1978) Recent developments in forensic anthropology. *Yearbook of Physical Anthropology* **21**:160–173.

Krogman WM (1939) A guide to the identification of human skeletal material. *FBI Law Enforcement Bulletin* **8**:1–29.

Krogman WM (1962) *The Human Skeleton in Forensic Medicine*. Springfield, Illinois: Charles C. Thomas.

Krogman WM and İşcan MY (1986) *The Human Skeleton in Forensic Medicine*, 2nd edn. Springfield: Charles C Thomas.

Maples WR and Browning M (1994) *Dead Men Do Tell Tales*. New York: Doubleday.

Reichs KJ (1998) Forensic anthropology: A decade of progress. In: Reichs KJ (ed.) *Forensic Osteology*, 2nd edn, pp. 13–38. Springfield, IL: Charles C Thomas.

Rhine S (1998) *Bone Voyage: A Journey in Forensic Anthropology*. Albuquerque: University of New Mexico Press.

Snow CC (1982) Forensic anthropology. *Annual Review of Anthropology* **11**:97–131.

Stewart TD (1979) *Essentials of Forensic Anthropology*. Springfield, IL: Charles C Thomas.

Ubelaker DH and Schammel H (1992) *Bones: A Forensic Detective's Handbook*. New York: Harper Collins.

Wilder HH and Wentworth B (1918) *Personal Identification: Methods for the Identification of Individuals, Living and Dead*. Boston: Gorman.

Forensic Medicine

W Bonte, Heinrich Heine University, Duesseldorf, Germany

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0565

Introduction

In the mid-1960s Robert P Brittain published a series of papers on the origin of legal medicine. When

reporting on the famous criminal code of Emperor Charles V of Germany, the *Constitutio Criminalis Carolina*, promulgated in 1532, he wrote that 'the code, important in the history of law generally, is a landmark of the first importance in the history of legal medicine' and continued:

> 'It has commonly been considered as the true start of legal medicine and hence Germany has been hailed as the country which gave birth to the discipline. It has been said that it caused medical men to be called on in legal matters for the first time. This is not strictly true. They had been called on before as earlier enactments show. Without in any way minimizing the advance which the Carolina represented, it would be wrong to consider it as a phenomenon which occurred without logical antecedents, and as implying that legal medicine arose by a kind of spontaneous generation.

Nearly the same had already been stated by Mende in 1819, when he wrote:

> The Carolina is praised as the true mother of forensic medicine. But if one believes, as some do, that these orders had no connection with the ideas of that time and emerged completely from nothing, and if one intends to derive forensic medicine as an up to then totally unknown discipline from the Carolina, he will be heavily mistaken.

One must agree. There are older roots, even very old ones.

The Ancient World

There are indications that medicolegal matters were already being dealt with in ancient Egyptian and Babylonian times. Sidney Smith expressed the opinion that Imhotep, the adviser of Pharaoh Djoser (2720–2700 BC), priest and physician, was familiar with matters we would call medicolegal. It appears that in his era the examination of dead bodies was already customary. According to Dérobert some paragraphs of the code of Hammurabi (1728–1686 BC) deal with medical malpractice. However, there is no proof that a medical speciality worthy of being called forensic medicine really existed.

The same goes for the ancient Greek and Roman worlds. The progenitor of the medical profession, the Greek Hippocrates (460–370 BC), *inter alia*, commented on the lethality of wounds. The first Roman law which dealt, among others, with medicolegal questions was the *lex duodecim tabularum* of 449 BC. It is well documented that Antistius, personal physician to Caesar, examined his corpse after he was assassinated in 44 BC and found that among 23 stab wounds only one chest wound was lethal. The famous Greek-Roman physician Galenus (AD 131–201) wrote a treatise on forensic–psychiatric problems. However, it appears that the need of a medical expert for a court decision was not recognized at that time.

The only concrete indication that medical experts could be asked to assist in judicial decisions was found in the *corpus iuris civilis* of the East-Roman emperor Justinian (AD 482–565): 'Medici non sunt proprie testes, sed majus est iudicium quam testimonium'. (Physicians are not ordinary witnesses, but give judgment rather than testimony.) Sidney Smith derived from this text: 'The Justinian enactments may be held to represent the highest point of achievement in the way of defined forensic medicine in the ancient world.' It appears that the highest point at the same time marked the end of this achievement under Roman influence.

Early Documents from Outside Europe

In India, the *Artha Shastra* of Kantilya was the law code of the fourth to third centuries BC. In cases of unnatural death there was an examination of the corpse. *Shusruta Samhita*, written between AD 200 and 300 by Shusrut, the father of Indian surgery, described forensic medicine in detail. This book also covers toxicology. Similar texts can be found in the *Baglebata* and the *Madhabakar* in the seventh century. There is, however, no clear indication that medical witnesses played a role.

In China, in the *Feng Zhen Shi* (around 300 BC), a number of cases of medicolegal examination were recorded. From the tenth century, several textbooks on medicolegal matters came down to us written, *inter alia*, by Djen Sin and by He Nin and his son He Min. In 1247 the famous textbook *Xi Yuan Ji Lu* (The Washing Away of Wrongs) was edited by the forensic scientist Song Ci (1186–1249). It was translated into 19 languages and republished with amendments up to the end of the nineteenth century.

In Egypt, Abu Balcer Zakaria El-Razi (865–925) edited the *El-Hawi* (Encyclopedia), which, among others, dealt with medicolegal problems. It was translated into Latin and is said to have been one of the nine volumes constituting the whole medical library of Paris University in 1395. In 1036, lbn Sina published *Kitab El-Shiffa* (The Book of Healing) and a couple of years later his second book *Ghanone Fi El-Tib* (The Book of the Legal Principles of Medicine). The latter was also translated and used as a textbook in many western universities, particularly in France and Italy.

Germanic Laws

Medieval courts were an outgrowth of the tribal courts of the Germanic peoples. The tribal judges

supervised the proceedings and executed the judgments rendered by the popular assemblies which met regularly throughout the year. During the development of Germanic tribal organization into territorial states, the primitive tribal courts underwent a corresponding evolution, increasing in number and becoming differentiated.

It was those Germanic peoples conquering the Roman Empire, whom the Romans called Barbarians, who for the first time laid down in written laws that courts, when evaluating wounds, had to rely on medical expertise. These so-called *Leges Barbarorum*, drawn up between the fifth and ninth centuries, have so many similarities with each other that it is likely that the basic legal concepts pre-existed among the Germanic tribes and were passed on by word of mouth.

They were codified in the sequence in which the single tribes entered Roman soil: the laws of the West Goths (*lex Euricianus*), written down under King Eurich in 475–476, which were renewed in the sixth century (*lex Visigothorum*); the laws of the Burgundians (*lex Burgundionum*), codified in 480 under Gundobad; the laws of the Salian Franks (*pactus legis Salicae*) in 507 and of the Ribuarian Franks (*pactus legis Ribvariae*) during the sixth century; of the Langobards (*edictus Rothari*) in 643; of the Alemanni (*pactus legis Alamannorum*) during the first part of the seventh century, and eventually, during the eighth century, by order of Charlemagne (742–814), the laws of the Bavarians, Saxonians, Thuringians, Chamavian Franks and Frisians.

All these laws contain extensive compilations of different forms and localizations of wounds and compensations (‘Wergeld’) which had to be paid by the perpetrator. From specific phrases it can be derived that medical experts not only were required for the assessment of concrete findings but were already involved in the formulation of the wording of the laws. It is interesting that, in the Visigothic code, the physician was entitled to a statutory fee of 12 solidi for each medical student he instructed. This demonstrates that medical schools already existed at that early time.

At the beginning of the ninth century, Charlemagne established *capitularies*, which were a compilation of Germanic laws and of ancient Merovingian codes. They included instructions to the judges that they must seek the support of medical evidence and rely on the advice of physicians, especially in questions of blows and wounds, of infanticide and suicide, of rape, of bestiality and of divorce on the grounds of impotence.

Of these laws, only the *lex Visigothorum* survived until the beginning of the modern era. After the end of the *reconquista* in the late fifteenth century, it was even translated into Spanish and, with the name of *Fuero Juzgo*, came into force again. Many of the other laws were also renewed at later times, but no revised versions appeared after the tenth century among the south and middle Germanic tribes. But obviously they were not forgotten and survived as established rites among the peoples. This can be derived from many observations. One is that the founder of the law school that in 1119 became the University of Bologna, the Langobard Irnerus, during the last part of the eleventh century, taught Roman as well as Langobardic law.

The development in England differed somewhat from that in continental Europe. The Germanic tribes of the Angles, Saxons and Jutes, which settled in England during the fifth and sixth centuries, had laws similar to those of the continental Saxons, Frisians and Thuringians. However, in contrast to the *leges Barbarorum* they were written in the native language. After the separation from the continent, a clear development towards feudal principles occurred. The former popular assemblies were replaced by a Council of the Great (*witenagemot*) and subordinate county courts (*scirgerefa*, origin of sheriff). There was an antithesis of common laws to privileges.

Whereas in continental Europe physicians or surgeons were responsible for the medicolegal examination of wounds and corpses, in England it was the coroner, a layman whose duty it was to hold inquests over dead bodies. The first indication of the existence of a coroner system in England can be found in the *Charts of Privileges* (925), which make reference to a grant of the coroner’s office by King Athelstane to an English noble. Athelstane of Wessex was the king who united the Anglo-Saxons during his reign (924–939).

Medieval Town Charters and Related Laws

In the year 1100 the so-called Assizes of Jerusalem were promulgated, a code of laws framed for the kingdom of Jerusalem at the instance of Godfrey de Bouillon, the crusader. The code determined, *inter alia*, that if, because of alleged illness, a vassal could not appear before the lord’s court to plead his case, ‘the Lord must send to this man’s house three of his men to decide the issue, a physician and an apothecary and a surgeon’. In cases of murder, these three experts were also sent, ‘and they must say what is the matter with him, and where he has been injured, and with what instrument it seems to them that the injuries have been inflicted’. As the laws and customs of the kingdom of Jerusalem were derived from those already existing in Europe, it is more than probable

that such medicolegal examinations were of common occurrence in Europe itself at this period. Similar regulations existed at the same time in Antioch: knights could only excuse their nonattendance before the court when medical experts confirmed the alleged illness.

The basic legal ideas of the Germanic laws were also reflected by the edict (1154) of the Norman king, Roger II of Sicily, and another edict (1221) of his grandson, the Hohenstaufen emperor, Frederick II (who was also the grandson of Frederick Barbarossa), which both provided for the appointment of physicians to help the courts. These edicts had a strong influence on the Italian town charters between the thirteenth and sixteenth centuries, which contributed much to the development of a systematically consolidated medicolegal discipline.

The Italian town charters without doubt were codified with the help of the faculties of law of the newly established universities (Parma 1066, Bologna 1119, Modena 1175, Perugia 1200, Padua 1222, Naples 1224, Siena 1240). In 1249 the physician Hugo de Lucca was appointed sworn expert of the magistrate of Bologna. One of his reports of the same year (on a case of abortion) still exists. It is likely that he was the first to perform legal autopsies, between 1266 and 1275. However, the first documented legal autopsy report of Bologna, in a case of suspected poisoning, was signed by Bartolomeo da Virignana in 1302. The Faculty of Medicine of the University of Bologna was founded in 1156, but only in 1306 did it become independent of the Faculty of Law. This means that these first significant developments must have taken place under the supervision of the Faculty of Law.

Most of the Italian town charters determined that two experts, generally a physician and a surgeon, were responsible for postmortem examinations (e.g. Padua 1315, Mirandola 1386). The town charter of Bologna (1288) stipulated that a medical expert must be at least 40 years of age and a citizen of Bologna for at least 10 years. The town charter of Pavia (1394) even regulated the performance of legal autopsies. Several reports on postmortem examinations, not always autopsies, still exist (e.g. Bologna 1289, Verona 1450, Brescia 1470, Milan 1502).

Similar regulations can be found in the canon laws. The decrees of Pope Innocent III (1209) provided for the consultation of medical experts in cases of injury and intoxication, among others. External examinations of all corpses were ordered. The *Nova compilatio decretarium* of Pope Gregory IX (1234) dealt, *inter alia*, with matters in which, to bring evidence against the accused, a physician had to be called upon.

In France, the Bishops of Maine and of Anjou had medical experts in their service from the eleventh century, and there were surgeon-experts in Paris from the twelfth century. Philip the Bold, in 1278, provided for sworn surgeons in legal matters. Examples of medical reports exist from 1330 onwards. As in Italy, in France the first universities were founded around the same time (Paris 1211, Montpellier 1289). In 1374 the right of autopsy was given to the Faculty of Medicine of the University of Montpellier by the Pope.

A similar development occurred in Spain, where the first university in Salamanca was founded in 1218. Juan de Castillo (1406–1454) wrote a *Cedula Real* (Royal Decree), which organized the *Protomedicato* (Court of Justice for Medical Persons). The *Protomedicato*, with the help of court physicians, was responsible for the sentencing of physicians and surgeons in cases of medical malpractice. During the seventeenth and eighteenth centuries the principle of the *Protomedicato* was exported to Latin America.

In Germany, from the eleventh century the towns achieved independent jurisdictions. Written laws became necessary and were modeled on the Germanic common laws. The similarities in the judgment of injuries between the Germanic laws of the fifth to eighth centuries and these town charters are surprising. The laws included the same catalogs of wounds, arranged by severity and location, and compensation to be paid; even similar wordings can be found.

One of the older laws which has come down to us is the *Sächsisches Weichbildrecht* (Saxonian settlement law, 1237–1250), which was derived from the older *Sachsenspiegel* (Saxonian mirror, 1225). This written law provided for medical experts who had sworn an oath to assist the courts. Similar regulations can be found in the town charters of Lubeck (1224) and Hamburg (1270). These charters were adopted by several other north German and northeastern European towns, reaching as far as Reval (1282). Similarities can also be found in the town charters of Nuremberg (1350) and Constance (1387). The town charters of Goslar (early fourteenth century) and of Magdeburg (late fourteenth century) provided for sworn medical experts in all cases of injury and of murder. The experts were regarded as helpers of the courts, not as witnesses. Written expert reports of the fourteenth century still exist in Magdeburg, Stendal and Goslar.

As in Italy, this development coincided with the establishment of universities (Prague 1348, Vienna 1365, Heidelberg 1386, Cologne 1388, Leipzig 1409). Their medical faculties discussed and criticized court decisions, especially in capital offenses. As far as is known, the first opinion was given by the medical faculty of the University of Cologne in 1478. It included the following advice: ‘It is useful and necessary

that those who die unexpectedly – God forbid this, but unfortunately it happens so often – are opened and dissected immediately in order to examine the organs and find the cause of death or the lethal disease.' A similar opinion was given by the Leipzig faculty in 1517. These opinions doubtlessly instilled a proper caution in the medical experts, who knew that their conclusions would be the subject of scrutiny. From this time on, forensic medicine shifted more and more from a merely practical to a scientific discipline.

If one compares the medicolegal aspects of the Italian and German town charters, it is remarkable that the same or very similar laws were in existence at about the same time. To give an example: both the Italian and the German town charters set great store by the exact diagnosis of the number of lethal wounds. This number limited the number of possible perpetrators and, accordingly, the number of those who could be charged with the crime, a concept that was later abolished.

Developments in England differed completely from those in continental Europe. When the Normans conquered England in 1066, they imposed the Carolingian judicial system on the Anglo-Saxons. A long struggle between King and landed nobility was won by the Crown and from the thirteenth century English courts became organized on a centralized basis. During the reign of King Henry II (1154–1189), the highest court, filled with professional judges, developed the principle of common law. Common law in a sense is unwritten law. It was defined as the customary law of England, declared and expounded by the judges as cases of dispute arose. The principles of such court decisions (precedents) were binding on later decisions. Common law thus contrasted with the statute law of the continent.

The office of the coroner was formally described in 1194, when the justices were required to provide that three knights and one clerk were elected in every county as 'keepers of the pleas of the crown'. The term 'coroner' is obviously a corruption of crowner. The coroner had the duty to hold inquests over dead bodies. He also had to inspect wounds of living individuals, to record the accusation against another individual and, if the wounds appeared likely to be fatal, to arrest the accused individual. The coroner was neither physician nor surgeon, and there are no indications that he was assisted by medical experts.

The Caroline Code and Developments in Continental Europe

There is no doubt that the German town charters had a strong influence on the codification of the Bamberg code (1507) and the Caroline code (1532). Although the *Constitutio Criminalis Carolina* followed the Bamberg code and was clearly based on it, it was the *Carolina* that became famous and is today accepted as a landmark in the history of forensic medicine. The main reason is that the Emperor Charles V of Germany, who promulgated his criminal code in 1532, insured that it was to govern his empire, which extended over much of Europe. Thus it provided a uniform system of German penal jurisprudence for a great part of Europe. The Caroline code took into account the fact that many questions could not be answered purely by legal methods; it obliged judges to take formal evidence from physicians, sworn in advance, in cases of doubt or difficulty where death had resulted from violence, whether criminal or accidental.

Importantly, this code introduced the practice of conducting medicolegal autopsies. Although more than a century passed before they became generally obligatory, instead of just being permitted and used only occasionally, as was previously the practice, the Caroline code thus opened the way for forensic medicine to develop as a separate discipline, as indeed it did during the succeeding hundred years.

The first step in this direction was the publication of monographs and systematic treatises. The earliest and most significant authors were Ambroise Paré (France 1575; Paré was trained in Bologna and Montpellier), Juan Fragoso (Spain 1581), Fortunatus Fidelis (Palermo 1597), Giovanni Battista Codronchi (Italy 1597), Felix Platter (Switzerland 1614), Paolo Zacchia (Italy 1621), Bernardus Suevus (Germany 1629), Melchior Sebitz (Germany 1638), Gottfried Welsch (Germany 1660) and Johannes Bohn (Germany 1689).

The next logical step in evolution was that forensic medicine became the subject of special instruction, which occurred in the seventeenth century. The first lecture was held in 1650 by Michaelis in Leipzig. Other universities followed at the beginning of the eighteenth century (in Germany these were Giessen 1700, Halle 1703, Erfurt 1718, Jena 1719, Berlin in 1724). The first medicolegal journal appeared already in 1782 in Berlin, published by Uden and Pyl. From the end of the eighteenth century, chairs of forensic medicine were created in the medical faculties of many German universities (Heidelberg 1762, Vienna 1805, Prague 1807, Berlin 1820). Eventually institutes of forensic medicine were founded (the first were Vienna 1818 and Prague 1820). In 1905 the Deutsche Gesellschaft für gerichtliche Medizin (German Society of Forensic Medicine) was established.

Although the traditional unity of the discipline remained unchanged, some remarkable new developments occurred. Criminal investigation institutes

were founded by the police, which little by little took over parts of the work previously carried out by the university institutes, such as ballistics, fingerprints, traces. These institutes are well respected today and are also well staffed and well equipped. A second change emerged from rapid developments in the natural sciences. During the eighteenth and nineteenth centuries forensic toxicology was mostly in the hands of physicians. It became necessary to attract chemists to the discipline, and, with the emergence of forensic serology, biologists as well. They are now integrated in institutes of forensic medicine, which makes good sense.

However, the typical peculiarities derived from the historic development in Germany, as in other continental states, remained unchanged. Legal medicine is based at the university and is thus independent of judical, investigative and political authorities. It encompasses most subdisciplines of the forensic sciences: forensic pathology and clinical forensic medicine; clinical and forensic toxicology; forensic serology (including biological traces and paternity testing); insurance medicine; traffic medicine (including forensic alcohology); and medical law and ethics. In some institutes the following subdisciplines can also be found: forensic psychiatry; forensic odontology; forensic anthropology; and forensic criminalistics (hairs and fibers, toolmarks and impressions, shot-distance determination).

Developments in the English-speaking World

England did not join the rapid progress of the continent. There was little development in the coroner system until the middle of the nineteenth century. In 1860 the fee system was abolished and salaries were established for the county coroners. In 1888 the election of the coroners by freeholders was abolished and an appointee system was developed, under which the head of the local govermental unit appointed the coroner. There remained, however, no minimum qualifications for office. These were established in 1926, when a law was enacted requiring 5 years experience as a medical practitioner, barrister or solicitor if the individual was to qualify as a coroner.

The impact of the development of institutes of forensic medicine in continental Europe was felt in England. In 1788 the first systematic book on medical jurisprudence in English was edited in London by Samuel Farr. However, the development of legal medicine was more satisfactory in Scotland than in England. In 1789, Andrew Duncan, Professor of Physiology at the University of Edinburgh, began giving lectures in legal medicine. He published a textbook in 1792. In 1807 the first chair of legal medicine in the English-speaking world was established at the University of Edinburgh and occupied by Andrew Duncan Jr. One of his successors, Robert Christison, published the first monograph on forensic toxicology in the English language in 1829. In Glasgow the first lectures on the topic were given in 1826 by James Arbour. A chair of medical jurisprudence was established in 1839. In Aberdeen, lectures on the subject were given by Francis Ogston from 1839. The university chair was established in 1857.

The most famous name in English legal medicine was that of Alfred Swaine Taylor. He became Professor of Medical Jurisprudence at Guy's Hospital Medical School in 1834. In 1836 he edited a famous textbook. At King's College a chair of medical jurisprudence was established in 1844. The first professor was WA Guy.

At the beginning of the twentieth century forensic medicine developed more or less independently of medical school teaching of the subject. This was related to the foundation of Scotland Yard, which became an important police investigative agency. Coroners experienced the advantages of calling upon Scotland Yard to assist in their investigations. In investigations into deaths, Scotland Yard detectives soon found themselves in need of assistance from pathologists and toxicologists; such experts were appointed and given the title of Home Office Pathologist or Home Office Analyst.

The development of the adversarial system of jurisprudence in England, however, also created problems for medicolegal experts. Francis Camps expressed these thus:

> History has revealed that the physician in the courtroom is the traditional contribution of legal medicine to justice. With the development of the common-law adversary system of jurisprudence in Britain, the medical expert became a partisan, or at least felt he became such. In the older continental jurisprudence the inquisitorial system utilized all witnesses, including the physician, as court (not party) witnesses. Perhaps this legal history alone explains why legal medicine has been more successfully used for justice in Europe than in the common-law countries, such as the United States, which have accepted Britain's legal system.

The early American colonists, originating in England, brought the coroner system with them in essentially the state of development it had reached by the early 1600s. There still exist documents of coroner's inquests in New Plymouth (1635) and in Maryland (1637). Early autopsies are recorded in Massachusetts (1647) and in Maryland (1665).

James S Stringham of New York was the first to

lecture on legal medicine in the US (1804) and was the first professor of the subject (1813). He came from Edinburgh and was influenced by the teaching of the Duncans there. His first textbook was published in 1814. Even earlier (1811), Benjamin Rush edited a textbook in Philadelphia. He was also a medical graduate of Edinburgh. In 1812–1813, Charles Caldwell gave a course of lectures on legal medicine at the University of Pennsylvania. TR Beck, pupil of Stringham, became Lecturer in Medical Jurisprudence at the College of Physicians and Surgeons in the Western District of the State of New York in 1815. Together with his brother, JB Beck, he published a famous textbook in 1823. Later on, in 1826, TR Beck was appointed professor of the subject at Fairfield College.

The Code of Public General Laws of Maryland (1860) authorized the coroner to require the attendance of a physician in cases of violent death. The 1868 legislature even authorized the governor to appoint a physician as sole coroner of the city of Baltimore. In Boston, in 1877, the coroner system was replaced by the medical examiner system. However, the medical examiners did not have the right to order autopsies. This was not corrected until the 1940s. No central laboratory for toxicological analyses was available. It was only in the 1980s that a true state medical examiner system was established in Massachusetts.

In New York, the medical examiner system was introduced in 1915. It was the famous Milton Helpern who brought together this office with an Institute of Forensic Medicine as a teaching arm of New York University. In 1937, the first intramural department of legal medicine in an American medical school was established at Harvard University.

It is remarkable that the Medico-Legal Society of the City and State of New York, founded in 1867, is the oldest scientific society on the subject in the world.

See also: **Anthropology**: Overview. **History**: Forensic Sciences; Crime-scene Sciences; Fingerprint Sciences; Forensic Anthropology in the USA. **Legal Aspects of Forensic Science**.

Further Reading

Brittain RP (1965) Origins of legal medicine: Constitutio Criminalis Carolina. *Medico-Legal Journal* 33:124–127.

Brittain RP (1965) Origins of legal medicine: the origin of legal medicine in Italy. *Medico-Legal Journal* 33:168–173.

Brittain RP (1966) Origins of legal medicine: Leges Barbarorum. *Medico-Legal Journal* 34:21–23.

Brittain RP (1966) Origins of legal medicine: the Assizes of Jerusalem. *Medico-Legal Journal* 34:72–73.

Brittain RP (1966) Origins of legal medicine: the origin of legal medicine in France. *Medico-Legal Journal* 34:168–174.

Brittain RP (1967) Origins of legal medicine: Roman law: Lex Duodecim Tabularum. *Medico-Legal Journal* 35:71–72.

Camps FE (1968) *Gradwohl's Legal Medicine*, 2nd edn, pp. 1–14. Bristol: Wright.

Dérobert L (1973) Histoire de la médecine légale. *Zacchia* 48:1–37, 161–192, 341–382, 533–547.

Mallach HJ (1996) *Geschichte der Gerichtlichen Medizin im deutschsprachigen Raum*. Luebeck: Schmidt-Roemhild.

Mende LJC (1819) *Ausfuehrliches Handbuch der gerichtlichen Medizin für Gesetzgeber Rechtsgelehrte, Aerzte und Wundaerzte*, vol. I. Leipzig: In-der-Dyk'sche Buchhandlung.

Schmidt O (1953) Gerichtliche Medizin in den ersten geschriebenen Rechten germanischer Staemme. *Deutsche Zeitschrift fuer die Gesamte Gerichtliche Medizin* 42:121–132.

Smith S (1951) The history and development of forensic medicine. *British Medical Journal* 1:599–607.

von Neureiter F (1935) Anfaenge gerichtlicher Medizin nach den Stadtrechten des deutschen Mittelalters. *Deutsche Zeitschrift fuer die Gesamte Gerichtliche Medizin* 24:1–7.

Forensic Sciences

D Wielbo, Center for Environmental and Human Toxicology, University of Florida, Gainesville, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0564

Introduction

Forensic science is the application of science pertaining to the law and requires the complementary interaction of a wide range of scientific specialties and disciplines. The first aspect of applied forensic science begins with the identification, individualization or classification of physical evidence. For some types of evidence, identification or individualization may only be possible after conducting chemical or scientific tests. Types of evidence that require testing to insure accurate identification include blood stains, body fluids, drugs, arson accelerants and other chemicals. The identification of unknown substances or

objects may be achieved by comparing their characteristics with those of known standards, previously established criteria or database information. In the forensic examination of fibers and hairs, determination of fiber type, form, dye composition, elucidation of color, species or anatomical origins utilize class characteristics for such identification. The ultimate goal of the identification process in forensic science is individualization: to say that a particular piece of evidence originates from a specific locus, scene or person. In reality, few types of evidence can be unequivocally individualized like fingerprint and DNA evidence. All other types of evidence, if appropriately identified, can be said to be consistent with originating from a particular source, site or individual. Many of the classical databases, techniques and tests now routinely used in forensic identification have been constructed, developed and refined in a process which, for some disciplines, has taken centuries, and for more recent technologies, decades. To depict the historical development and growth of forensic sciences, this narrative will broadly address documented instances of the organizational and developmental aspects of several disciplines and specialties within the profession, outlining the progression from basic conception to the application of some methodologies and techniques used today. The evolution of the field originates with the worldwide development of a number of different specialty or subspecialty scientific disciplines, with the rate of advance or development of each discipline being extremely variable, due to complexity, geographical location or availability of financial resources and technology.

Forensic Medicine and Toxicology

The earliest documented forensic specialty could probably be considered to be forensic medicine, also known as legal medicine or medical jurisprudence, and is the application of medicine and medical science to answer legal problems. Records indicate that the first documented dissertation on forensic medicine was written in China in the sixth century by an individual named Hsu Chich'Ts' si. Although this work was apparently lost, a second Chinese manuscript still exists. Completed in 1247, the *Hsi Duan Yu* (The Washing Away of Wrongs) provides what is thought to be the first alliance of medicine and law and supplies specifications on distinguishing death by drowning or strangulation from death by natural causes. Further records suggest that legal medicine really began to prosper sometime in the sixth century, heralding the development of modern day science and medicine. It was not until around the end of the eighteenth century that the first appearance of legal medicine experts in the courtroom was documented.

The development of modern-day chemistry is considered to have begun at the end of the eighteenth century, paving the way for the development of modern toxicology. Closely related to forensic medicine, forensic toxicology centers on the determination of toxic substances in human tissues and organs and the subsequent determination of the role any toxic agents may have in contributing to or causing death. In general, forensic toxicology cases entail some form of drug or alcohol abuse. One of the most influential people in the development of toxicology was Mathieu Orfila. Orfila moved to France in 1807 and eventually became dean of the medical school in Paris. He is considered to be one of the first experts to provide reliable scientific evidence in a criminal trial, while his academic studies provided toxicology a firm niche within the developing sciences. Orfila and associates are also credited with developing the first chemical test to detect arsenic, the poison of choice at that time. However, it was James Marsh, a Scottish chemist, who was considered to be the first to provide toxicological evidence supporting arsenic detection for a legal trial, around 1836. The Lafarge case, also conducted by Orfila, is documented as the first case in which the defense called an opposing expert, Francois Vincent Raspail, in an attempt to refute the scientific evidence of the prosecution's expert witness.

Personal Identification

Personal identification plays a mammoth role in forensic and criminal investigations, and many subspecialties in personal identification are still being developed and more frequently utilized. Forensic odontology uses dental records to facilitate human identification. This type of identification process is often called for in the identification of unrecognizable bodies, after mass disasters, and in the identification and comparison of bite marks. Also facilitating personal identification is the use of forensic anthropology. This discipline, mainly concerned with bodily or skeletal remains, has, over recent years, constructed a number of databases cataloguing differences in physical structure of the body as a function of sex and race. It was Alphonse Bertillon who first developed the most significant and earliest personal identification system, at the end of the nineteenth century. Bertillon's system of anthropometry, which involved a series of body and facial measurements, was developed to establish a process for use in the identification of habitual criminals. The system, also known as bertillonage, was devised in Paris and applied worldwide. Nowadays, it is considered that this system of identification cannot be utilized to uniquely identify one

person as being distinct from all others. Although the concept of anthropometry was considered theoretically possible, in practice, inconsistency in measurement techniques decreased the discriminating ability of the system. Despite this, Bertillon's system is still considered to have made a significant contribution to the development of forensic science, representing one of the first examples of the use of individualization for criminal investigation. Anthropometric systems were replaced as the science of fingerprinting was developed and the value of fingerprinting in personal identification and individualization was acknowledged.

Fingerprints

Fingerprint examination is probably the most used of personal identification techniques. Much of the work involves the study and classification of fingerprints, the development of latent prints and the comparison of known and unknown fingerprints. Despite the fact that Bertillon opposed the introduction of fingerprinting, he was the first person in Europe to use latent prints to solve a criminal case, as well as being among the first to use systematic crime scene photography. The development of fingerprinting for personal identification was originally the initiative of William Herschel, a British civil servant residing in India in 1877, who first publicly proposed that fingerprints might be useful in the identification of criminals. However, it was Henry Faulds, a Scottish physician working in Japan around the same time, who recognized the value of latent prints from a crime scene. The use of fingerprints in identifying a criminal was subsequently published in the journal *Nature*, and Faulds continued on to use fingerprints to identify a Tokyo burglar in 1880. As a result of his own studies and fingerprint pattern types described by Francis Galton, Juan Vucetich, an Argentinean police official, set up his own fingerprint classification system in 1891. In 1892, Galton published *Fingerprints*, the first text on the use of fingerprints in personal identification and crime investigation, and devised the first scientific methods for the classification of fingerprint patterns. Several years later, Sir Edward Richard Henry, another British civil servant in India, continued to devise a fingerprint classification scheme, cataloguing sets of fingerprints that could be retrieved for identification purposes; he subsequently developed the fingerprint classification system that would replace anthropometry in Europe and North America. This system was implemented in Europe when Henry was appointed head of Scotland Yard in the early 1900s. At the same time, in the United States, DeForrest initiated the first systematic use of fingerprints in personal identification. More recently, the Federal Bureau of Investigation's (FBI's) Integrated Automated Fingerprint Identification System (IAFIS) was initiated as a means of providing identification services to US law enforcement communities.

Criminalistics

The discipline of criminalistics encompasses all areas of trace evidence, such as soil, glass, hairs, fibers, blood and other body fluids, including, saliva, sweat, semen and vitreous humor. Criminalistics also includes arson, explosives, drug identification and investigation, interpretation of pattern and imprint evidence, and is by far the broadest of the disciplines of forensic science.

As well as conducting toxicology studies, Orfila was also responsible for the inception of early studies on blood and semen identification. Some of Orfila's earlier work laid the foundation for the development of some of today's routine tests. As early as 1827, Orfila discovered that extracting intact spermatozoa from seminal stains was extremely difficult and developed a series of chemical tests for semen identification. Following subsequent studies, other scientists established that microscopical examination of semen stains provided the most useful forensic information. In 1891, Hans Gross of Austria first promoted the concept of criminalistics by applying scientific knowledge and methods to facilitate the interpretation and analysis of physical evidence for cime investigation. Gross, also a legal expert, saw the function of criminalistics as the complementary collaboration of various forensic specialists with appropriate scientific backgrounds for the examination and interpretation of physical evidence. Gross was also responsible for the emancipation of a journal dedicated to forensic sciences. In France, Victor Balthazard and Edmond Locard conducted similar work. Balthazard, medical examiner for the city of Paris, also performed studies on probability models for fingerprints, bullet comparison, animal hairs and blood spatter patterns. In 1910, Locard set up the first police crime laboratory in Europe, in Lyon, the location of today's Interpol, and is recognized for the inception of the Locard exchange principle: 'Every contact leaves a trace', the basic tenet of forensic science. Also during the 1920s, Luke May was the first to develop and apply striation analysis for toolmark comparison, incorporating statistical analysis into his methodology. In the latter part of the 1920s, Los Angeles Chief of Police August Vollmer, of the Los Angeles County Sheriff's department, developed the first United States police crime laboratory. The FBI crime laboratory was subsequently established in 1932, and 5 years later Paul Kirk set up the first academic criminalistics program

in the United States at the University of California. In 1950, the American Academy of Forensic Science was formed, with the subsequent institution of the *Journal of Forensic Science*.

Ballistics and firearm identification

The specialization of ballistics and firearm examination relates to firearm identification, comparison of bullet markings and rifling, the identification of projectiles, cartridge and shell cases, and determination of bullet trajectories and damage. The historical development of these specialties begins around 1835 when Henry Goddard, working for Scotland Yard, first used bullet comparisons to provide evidence leading to the arrest of a murderer. The comparison revealed a flaw in the bullet that could be traced back to a flaw in the original bullet mold. Nearly 55 years later, Dr Alexandre Lacassagne, professor of forensic medicine at the University of Lyon, was the first to associate bullets with a specific gun, by comparing striations etched on the bullet from a murder victim and striations from the gun barrel from which the bullet was fired. Lacassagne was also one of the first to study the nature of blood spatter patterns. The discipline further developed around 1900, when Paul Jesrich took photomicrographs of bullets for comparison, and illustrated the use of this technique for the possible individualization of minutiae. It was not until 1913 that Victor Balthazard published the first scientific article on the significance of bullet markings and their relation to bullet individualization. Further developments occurred during the 1920s, as Charles Waite first cataloged information regarding weapon manufacture, and, together with Goddard, Gravelle and Fisher, further developed the technique of comparison microscopy for bullet comparison.

Questioned documents

Another division of criminalistics is questioned document examination, which entails the comparison and interpretation of handwriting; materials generated from typing, printing, facsimiles and photocopying; and the analysis and aging of paper, inks and materials used to produce documents. Such techniques were used recently to assist in the invalidation of the Hitler Diaries. Francois Demelle in France reportedly published the first study on questioned document examination in 1609, but it was not until the 1800s that document examination truly began in Europe. Photographers at that time attempted to broaden their professional expertise by providing services that incorporated document comparison, but highly publicized mistakes delayed the acceptance of the profession. One such mistake involved Alphonse Bertillon, previously acclaimed for his invention and application of anthropometry to personal identification. In the Dreyfus case, Bertillon testified that the accused had written a document that was the basis for the charge of treason. The accused's innocence was later proven. Old English laws that stated that writings were inadmissible as standards for comparison unless the writings were in evidence in a prior case, also allegedly hindered the lack of initial acceptance of the questioned document profession in the United States. The acceptance of questioned document testimony in courts eventually became a reality in 1913 via the enactment of Section 1731, Title 28 US Code. The code states that: 'the admitted or proved handwriting of any person shall be admissible for purposes of comparison to determine genuineness of other handwriting attributed to such person.'

Nowadays, the Questioned Documents Unit of the FBI crime laboratory examines all aspects of documentary evidence, including hand printing, typewriter ribbons, printers, watermarks, erasures, alterations, obliterations, safety paper, charred paper, graphic arts, plastic bags and product tampering. The unit maintains the database files, such as the Anonymous Letter File, Bank Robbery Note File, National Fraudulent Check File, Office Equipment File and Watermark File.

Soil examination

The main contributor to the development of forensic soil and particulate examination was, again, Edmond Locard. In 1929 he made the observation that it was almost impossible for anyone to participate in any activity without removing soil or dust particles from the site, as a result of contamination of the person's body, clothes, tools or vehicle. However, Locard credited Sherlock Holmes with the idea that botanical or earth materials could have important evidential value. Arthur Conan Doyle books, published between 1893 and 1897, suggest that soil collected on a person at a crime scene could be used as evidence to place that person at that location. The first recorded use of this type of evidence was in Germany in 1904, by Dr George Popp who developed and presented what is thought to be the first example of evidence in a criminal case using botanical materials. Throughout the 1920s Popp continued to pioneer the use of botanical identification in forensic work.

Identification and association of human hair

Rudolph Virchow, a professor and prosecutor in Berlin, Germany, who established that a questioned hair from a particular crime suspect was indistinguishable from that of the victim, reported the first forensic

investigation of human hair in 1861. By the early 1900s the significance of hair evidence in criminal investigations was routinely criticized and scrutinized by a number of medicolegal experts. In 1906, Hugo Marx, an official in the State Medical and Prison Medical Examiner's office in Berlin, wrote a dissertation on the value of forensic hair examination and its role in personal identification. A textbook published at the turn of the century, from the founder of the Vienna school of forensic medicine, also contained a chapter on the investigation of hair. In 1910 Victor Balthazard and his associate, Marcelle Lambert, produced *Le poil de l'homme et des animaux* (The Hair of Man and Animals). In this work they describe techniques closely resembling those still used in hair comparison today – mainly microscopical analysis to compare and observe morphological features. In 1931, John Glaister published a study of mammalian hairs and wool and their relevance to criminal investigation. The work contains approximately 1700 photomicrographs that show the structure of hair of humans and other mammals. Paul Kirk, criminalistics professor at the University of California at Berkeley, further developed hair analyses by improving the techniques of hair comparison, utilizing various physical and chemical properties in an attempt to facilitate the systematic individualization of human hair. Despite, cutting-edge advances in various other specialties of forensic science, the methods, philosophy and significance of forensic hair examination have not changed greatly since the early twentieth century. The Trace Evidence Unit of the FBI today holds extensive reference collections of human and animal hairs, natural and manmade fibers, feathers, ropes, cordage, woods and seeds.

DNA profiling

Evolving from classical serology, DNA profiling could be considered the modern-day technique revolutionizing personal identification in forensic science. In the mid-1980s Sir Alex Jeffreys developed the techniques allowing the profile analysis of DNA. After publishing his achievements in *Nature* in 1985, Jeffreys was subsequently called upon to apply his techniques to solve the first crime in 1986. In combination with the British Home Office Forensic Science Service, his DNA profiling techniques were used to identify Colin Pitchfork as the murderer of Dawn Ashworth and Lynda Mann in Leicestershire, England. Cetus Corporation furthered the developments of DNA profiling and molecular biology techniques in personal identification during the rest of the 1980s with the development of the polymerase chain reaction. In 1987, not only was DNA profiling introduced for the first time to US criminal courts, but the admissibility of DNA evidence was also challenged, resulting in the development and implementation of appropriate accreditation, standardization and quality controls for DNA and forensic laboratories. During 1996, the FBI DNA Analysis Unit began using mitochondrial DNA analysis. This type of analysis can be applied to small or degraded quantities of DNA from hair, bones, teeth and body fluids, allowing the examination of evidence that may not have been suitable for comparison prior to the development of this technique.

Accreditation

Finally, over the last 30 years or so, several organizations have emerged that have developed guidelines and regulations assuring forensic laboratory standardization, accreditation and certification as a guide to legal and forensic communities. Of the numerous organizations that exist, three have made significant impacts on the professional appearance of the field of forensic science. The American Society for Testing and Materials (ASTM) committee E-30 on Forensic Sciences was created in 1970, with the purpose of standardizing methods and terminology unique to the field. To keep pace with the rapid technological advances within the field, standards are continuously updated. Established in the mid 1970s, the American Society of Crime Laboratory Directors (ASCLD) is an international organization composed of crime laboratory directors whose mission is to promote 'excellence through leadership in forensic science management'. The ASCLD laboratory accreditation board (ASCLD/LAB) is a related organization that has published minimum standards for laboratories and their personnel to achieve, allowing formal accreditation by the organization. Around the same time the American Board of Criminalistics developed a series of examinations allowing the certification of individual forensic scientists in their particular area of expertise. Although individual certification processes are currently voluntary, it is probable that courts will come to expect such quality assurances as the number of certified laboratories and individuals using standard techniques continues to grow.

See also: **Accreditation of Forensic Science Laboratories**. **Criminalistics**. **Deoxyribonucleic Acid**: Restriction Fragment Length Polymorphism. **Document Analysis**: Handwriting; Analytical Methods; Forgery/Counterfeits; Ink Analysis; Document Dating. **Explosives**: Mechanism of Explosion. **Fingerprints (Dactyloscopy)**: Visualization; Sequential Treatment and Enhancement;

Identification and Classification; Standards of Proof; Automated Methods, including Criminal Record Administration. **Firearms**: Weapons, Ammunitions and Penetration; Range; Residues; Laboratory Analysis. **Hair**: Overview; Identification of Human and Animal Hair. **History**: Forensic Medicine. **Identification/Individualization**: Overview and Meaning of ID.

Further Reading

Almirall JR and Furton KG (1995) The evolution, practice and future of forensic science in the administration of justice – the importance of standards in forensic science. *Standardization News* April 23 (4).

De Forest PR, Lee HC and Gaenssten RE (1983) *Forensic Science – An Introduction to Criminalistics*. New York: McGraw-Hill.

Else WM and Garrow JM (1934) The detection of crime. *Police Journal* **21**:195.

FBI science and technology web site. http://www.fbi.gov/scitech.htm

Jeffreys AJ, Wilson V and Thein SL (1985) Individual specific fingerprints of human DNA. *Nature* **316**:76–79.

Kind S and Overman M (1972) *Science Against Crime*. London: Aldus.

Kirk P (1974) *Crime Investigation*. Malabar, FL: Krieger.

Lane B (1992) *The Encyclopedia of Forensic Science*. London: Headline.

Peng Z and Pounder DJ (1998) Forensic medicine in China. *American Journal of Forensic Medicine and Pathology* **19**:368–371.

Rudin N Forensic Science Timeline. http://www.forensicdna.com/forensictimeline.htm

Saferstein R (1982) *Forensic Science Handbook*. Upper Saddle River, NJ: Prentice-Hall.

Homicide *see* **Crime-scene Investigation and Examination**: Collection and Chain of Evidence; Contamination; Criminal Analysis; Fingerprints; Packaging; Preservation; Recording; Recovery of Human Remains; Scene Analysis and Reconstruction; Suspicious Deaths.

Horses *see* **Toxicology**: Equine Drug Testing.

I

IDENTIFICATION/INDIVIDUALIZATION

Overview and Meaning of ID

C Champod, Forensic Science Service, London, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0568

The Identification Process

> The real aim of all forensic science is to establish individuality, or to approach it as closely as the present state of the science allows. Criminalistics is the science of individualization.

This citation of Paul Kirk shows how essential the identification process is to criminalistics.

The definition of identification in forensic science may differ largely from the one accepted in science, where the term 'identification' is simply used to describe the attribution of an object to a definite class. In criminalistics, the identification process seeks ultimately individualization. For forensic scientists, identifying an object means that it is possible to distinguish this object from all objects considered.

n the forensic literature, the problem of identity of source is often treated by reference to 'class' and 'individual' characteristics (**Table 1**). Comparisons that lead to agreement only in class characteristics (without significant differences) will end up with 'group identification' conclusions. Only when individual characteristics are present in conjunction with class characteristics can positive identification or individualization conclusions be drawn. The definitions of 'class' and 'individual' characteristics are only conventional ways of describing selectivity. We will see that the problem of inferring identity of source and its pertinence to criminalistics is more complex than a simple dichotomy between class and individual characteristics.

As illustrated in **Table 2**, we will distinguish the forensic fields that lead frequently to individualization and those leading rarely (in the present state of the art) to individualization, but more commonly to corroborative evidence of various strength.

Philosophically, identity of source cannot be known with certainty, and therefore must be inferred. As Kwan has demonstrated, the hypothetical-deductive method (assisted by methods of statistical inference) provides a reasonable explanation of how criminalists infer identity of source.

The identification process can be seen as a reduction process, from an initial population to a restricted class or, ultimately, to unity. The initial population constitutes control objects or persons, depending on

Table 2 Classification of forensic evidence types with respect to their identification capabilities

Individualization	*Corroborative evidence*
Fingerprints	Microtraces (glass, paint, hairs, fibers)
Footwear marks	Biological fluids (now mostly DNA evidence)
Earmarks	Drugs and toxicology
Tool marks and firearms	Explosives and fire residue analysis
Questioned documents	Soils

Table 1 Distinction between 'class' and 'individual' characteristics in some fields

Field	*'Class' characteristics*	*'Individual' characteristics*
Fingerprint identification	General pattern, ridge count, ridge tracing	Minutiae, pore structure, ridge structure
Footwear mark identification	General pattern, size, manufacturing characteristics	Cuts, accidental acquired characteristics, transient wear features
Bullet identification	Caliber, number of grooves/lands impressions, angles of grooves/lands impressions, width of grooves/lands	Grooves/lands impressions (striae)

the type of evidence. We have the combination of two factors:

- A relevant population of control persons or objects defined by its size (and/or other particularities). Put in another way, each member of this population of sources can be seen as a possible source.
- A reduction factor resulting from the combination of concordant characteristics of determined selectivity. In fact, the reduction is proportional to the rarity or random match occurrence of these observed characteristics in that population. As Kwan indicates: 'this is the sheer rarity of a feature that is important as rarity of that feature with respect to the set of suspected sources being considered. It is important to stress that rarity is relative to the situation at hand.'

With respect to the size of the relevant population, an 'open set' framework will be distinguished from a 'closed set' framework. The open set framework implies that the population at large is considered; meaning, for example, that all living persons on Earth or all produced objects on Earth are taken into consideration as potential sources. The closed set framework corresponds to a situation in which the number of control objects or persons is restricted to a specified set of suspected sources (for example, by taking into account other evidence available describing the putative sources).

To illustrate the identification process graphically, we will assume the following case. A mark is found in association with a crime. Following inquiry, a potential source is submitted to the laboratory for examination. The comparison shows that the recovered evidence and the control share some features. Moreover, there is no significant difference to be accounted for. Given this scenario, the identification process may be illustrated as shown in **Fig. 1**.

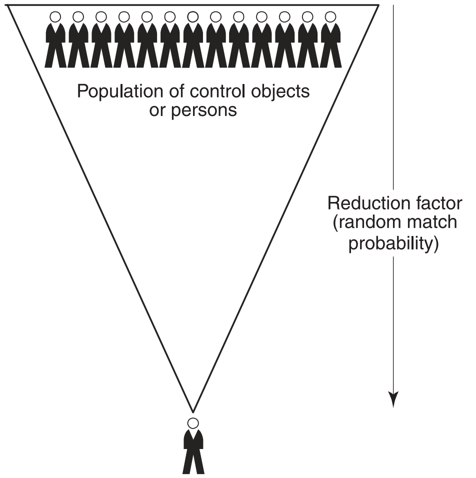


Figure 1 The identification process.

The identification process (in an open set or closed set framework) is a narrowing-down process, reducing the number of possible sources or hypotheses. The hypothesis that a designated suspect or object is the source, is proven by showing that all alternative hypotheses that could explain the phenomenon at hand are excluded. Also, to avoid error, the hypothetical-deductive method makes it imperative that all plausible hypotheses regarding possible sources are taken into account.

The importance of selecting features judiciously cannot be overemphasized. The criteria for selecting features fall into five areas (without taking cost into account): distinguishability, high intersource to intrasource variance, known variance in time, normalization (standardization) and independence. When individualization is the goal, the object must be defined by a unique set of properties (a set that no other source can share). Of course, each field is focused on different features and it is worth giving a few examples of adequately chosen features in some forensic areas (**Table 3**).

The Decision Schemes

In practice, for obscure reasons, the identification process leading to individualization is generally operated in an open set framework. This leads to two types of decision schemes: positive identification and corroborative evidence.

Positive identification or individualization

For Tuthill:

> The individualization of an impression is established by finding agreement of corresponding individual characteristics of such number and significance as to preclude the

Table 3 Features used in identification in some forensic fields

Field	*Features used to characterize (or ultimately to individualize)*
Fingerprints	General pattern, minutiae, pores and ridge structures
Footwear marks	Manufacturing characteristics (pattern, size, peculiarities of the manufacturing process) and acquired characteristics (wear features, cuts)
DNA	Various polymorphic loci on the DNA molecule
Microtraces	Optical (color, microscopic features, refractive index), physical (size, length, diameter) and chemical characteristics (Fourier-transformed infrared spectroscopy, elemental composition, pyrolysis coupled with gas chromatography, etc.)

possibility (or probability) of their having occurred by mere coincidence, and establishing that there are no differences that cannot be accounted for.'

Following this definition, the size of the control population is systematically set to its maximum (open set framework). This practice is generally used and required in fields like those of fingerprints, footwear marks, tool marks or firearms. For footwear marks, Bodziak stated that: 'The positive identification means that no other shoe in the world could have made that particular impression'. Analogous definition can be found for fingerprint identification: 'An "identification" is the determination that two corresponding areas of friction skin impressions originated from the same person to the exclusion of all others.'

The conclusion of a positive identification is then an opinion, a statement of probability expressing the view that the chance of observing on Earth another object or person presenting the same characteristics is zero. No contrary evidence will ever shake this certainty. According to Stoney, this decision is highly subjective, the identification process for fingerprint identification being described by him as follows:

> Beginning with a reference point in one pattern, a corresponding point in a second pattern is sought. From this initial point the examiner then seeks neighbouring details that correspond in their form, position and orientation. These features have an extreme variability, that is readily appreciated intuitively, and which becomes objectively obvious upon detailed study. When more and more corresponding features are found between two patterns, scientist and lay person alike become subjectively certain that the patterns could not possibly be duplicated by chance.
>
> What has happened here is somewhat analogous to a leap of faith. It is a jump, an extrapolation, based on the observation of highly variable traits among a few characteristics, and then considering the case of many characteristics. Duplication is inconceivable to the rational mind and we conclude that there is absolute identity. The leap, or extrapolation, occurs (in fingerprinting) without any statistical foundation, even for the initial process where the first few ridge details are compared.

The schematic description of the identification process can be refined to include the decision step for positive identification (individualization) (**Fig. 2**). This decision scheme calls for the following comments:

- The threshold, the leap of faith, is in essence a qualification of the acceptable level of reasonable doubt adopted by the expert. Jurists will interpret this threshold as an expression of the criminal standard 'beyond reasonable doubt' regarding the issue of the identification. Would jurists accept that the concept of reasonable doubt on the identification of a suspect is outside their province and that the threshold is imposed on to the court by the scientist? The response in the doctrine is negative, as expressed by the members of the Panel on Statistical Assessments as Evidence in Courts:

> the law may establish different thresholds for what is sufficient evidence in a case from those that statisticians would normally require in drawing conclusions. Clearly, the law must prevail and the statistician must adjust to the law's standards. Put another way, it is the utility function of the court that is appropriate, not the utility function of the statistician.

- It seems illegitimate to set the size of the relevant population at its maximum *a priori*. Indeed, the number of potential sources from which the mark could originate may be restricted by other evidence available (witness testimonies, other forensic evidence, etc.). Presenting the evidence in an open set framework is too conservative, adopting systematically the extreme defense attorney's position in trying to make the court believe that all persons or objects on earth could be the origin of the prints.

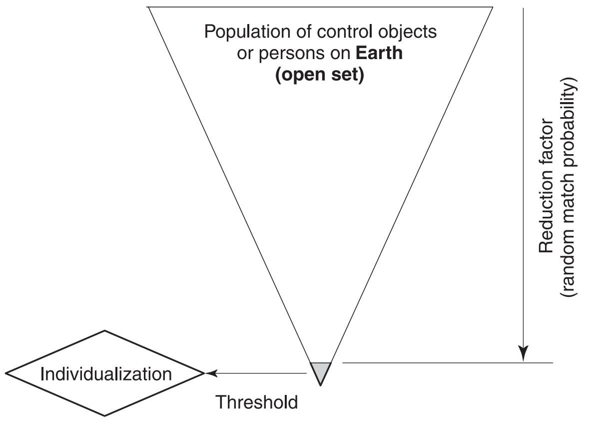


Figure 2 The identification process completed with a decision of positive identification.

Corroborative evidence

In some forensic fields (fingerprints, tool marks, footwear marks), practitioners have voluntarily excluded probability statements – other than exclusion and positive identification – from their conclusions. All items of evidence between these extremes are classified as 'inconclusive'. There is no logical reason for avoiding probability statements; the refusal of qualified opinions is a policy decision, even if the distinction of the arguments (policy or scientific argumentation) is not so clear in the literature.

(Appreciate the dogmatic statement proposed recently by a North American working group on fingerprint identification: 'Friction ridge identifications are absolute conclusions. Probable, possible, or likely identification are outside the acceptable limits of the science of friction ridge identification'.)

Indeed, the attention of jurists or scientists is too easily concentrated upon the ability of a technique to provide an absolute certainty. If the technique can positively conclude an identification, it is greeted as a panacea; if it cannot, it is damned as unreliable. This ignores the vital point that any technique will only function to a high degree of precision and accuracy under controlled conditions, and the conditions under which forensic scientists work are far from ideal. It follows that, in many cases, a forensic scientist will not be able to provide a definitive answer but only a probabilistic figure or opinion. If the ultimate set of specific features is not present or not detected in the evidence, then the examiner will not provide an identification but will express a probability statement, verbally or numerically, which attempts to assess the value of the evidence.

Each piece of evidence is relevant if it tends to make the matter that requires proof more or less probable than otherwise. Hence, a piece of evidence that only approaches the absolute identification constitutes relevant evidence that should not be ignored.

The identification process remains, in essence, a statistical process based on objective empirical data and/or on subjective evaluations related to the examiner's experience. When the evidence is only corroborative, it is necessary to form a conclusion in a way that reflects the statistical uncertainty.

In various fields which accept corroborative evidence, examiners have expressed the meanings of conclusion terms pragmatically. Their respective meaning relates to their power of reduction of the initial population. An agreement on the following terms seems to have been achieved: identification, very probable, probable, possible, inconclusive (could not be determined), appears not, negative, not suitable for analysis.

The identification process scheme can be refined as illustrated in **Fig. 3**. This decision scheme calls for two comments:

- The conversion between the objective statistical value and the verbal statement is never declared or explained. This naturally leads to obvious variations between examiners when assessing the same case.
- The blind and dogmatic allegiance to the open set framework is never questioned as for the positive identification.

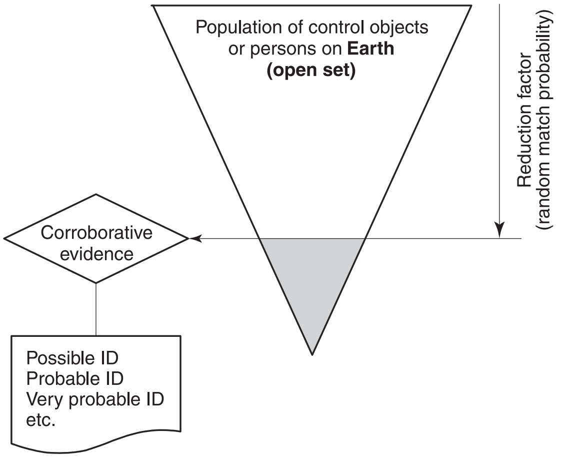


Figure 3 The identification (ID) process completed with corroborative evidence.

Identification and Statistics

Positive identification or individualization

From a statistical point of view, to conclude an individualization means that the probability (Pr) of the event, here the event 'identification', after examining the evidence is equal to 1. Because of the probabilistic nature of the identification process, the probability of 1 is an absolute that cannot ever be reached numerically. Hence, the examiner expresses an opinion of a positive identification when this probability is so near 1 that it can be reasonably set as 1. This probability of identification, given the evidence, can be represented in the form $\Pr(ID|E)$, where ID denotes the event identification, which is uncertain, and E denotes the information available, the evidence E, which has been taken into account. In this way, the vertical line can be seen as shorthand for the word 'given'. But we have seen that the identification process is related to the specificity of the print under examination, its random match occurrence in a population and the size of the suspect population being considered (N). The probability we are interested in is then conditioned on both E and N and becomes $\Pr(ID|E, N)$. Probability theory provides a tool to calculate $\Pr(ID|E, N)$. Let us denote the frequency of a random occurrence of the mark by f:

$$\Pr(ID|E,N) = \frac{1}{[1+(N-1)f]} \qquad \text{(Equation 1)}$$

$$f = \frac{1-\Pr(ID|E,N)}{\Pr(ID|E,N)(N-1)} \qquad \text{(Equation 2)}$$

If the population is set to $N = 5$ billion (including the control) and $f = 1$ in 5 billion, then, using Equation 1, $\Pr(ID|E, N)$ is close to 0.5. That means that if you

have a control object or person in hand with such a frequency, the probability that you have the wrong one is about 0.5! This expression is the correct statistical value even if it is counterintuitive. It has been largely demonstrated that intuition is a bad substitute for the laws of probability. If we accept that $\Pr(ID|E, N)$ must be above a certain threshold value in order to declare an identification, then it is possible to calculate the frequency that must be considered to achieve such a preset value. If $\Pr(ID|E, N)$ is fixed at 0.9998, $N = 5$ billion, using Equation 2 we see that f must be equal to 4.0×10^{-14}, which represents 1 in about 5000 times the size of the initial population of 5 billion. A conclusion of individualization (according to this threshold and in an open set framework) must be at least the expression of such a small random match probability.

Corroborative evidence

What is the statistical meaning, in the open set framework, of conclusions like: 'it is possible (or probable or highly likely) that this print/mark has been made by this particular item'? Such verbal scales can be translated into numbers expressing $\Pr(ID|E, N)$. It has been demonstrated that these terms are understood quite uniformly (in a numerical conversion) by experts or jurists. The numerical conversion is given in **Table 4** (columns 1 and 2). From these probabilities, their corresponding frequencies (f) can be derived, using Equation 2, considering an open set framework ($N = 5$ billion) as shown in Table 4 (column 3).

The analysis of Table 4 leads to a paradox. In the open set framework (the one claimed to be adopted in major forensic fields like footwear marks, firearms or tool marks), the rarity of the shared characteristics which corresponds to the most negative verbal statement is quite small (2.0×10^{-7}). For the court, every piece of evidence that would lead to such small frequencies will be considered as highly relevant (by analogy with DNA evidence, for example). Hence, the verbal statement proposed here does not make this evidential value very clear.

To escape this paradox, forensic scientists could be tempted to adjust the size of the relevant population; hence, to pass from an open set framework to a closed set framework. To understand if it is legitimate to do so, we will briefly introduce a proposal for an alternative inferential system. This proposal advocates the use of a bayesian model to assess the value of identification evidence.

The Bayesian Framework

In the bayesian framework, it is not assumed that the size of the relevant population is at its maximum (the world population of control objects). It seems that this relaxation constitutes the major difference between the two approaches (strict open set versus bayesian approach). Both closed set and open set situations can be handled in a bayesian framework; the open set constitutes a specific situation.

From information gathered through normal investigation or through the use of a database, a limited number of people or objects, or a single person or object, can be pinpointed in a general or limited population. There is no way to go back along the time axis and deduce the identification. The only way to evaluate the strength of the evidence is to consider this evidence from two perspectives: that the designated person or object is, or is not, the source.

The bayesian approach permits the revision of a measure of uncertainty about the truth or otherwise of an issue (here the identification ID) based on new information. It shows how data can be combined

Table 4 Relationship between verbal statements in qualified opinions and the probability (Pr) of identification, along with the corresponding frequencies f (assuming a relevant population of 5 billion)

Verbal statement	$\Pr(ID\|E, N)$ (%)	Frequency f
Positive terms		
Likelihood bordering to certainty	99.9	2.0×10^{-13}
Highly (very) likely	98	4.1×10^{-12}
Likely	85	3.5×10^{-11}
Very well possible (plausible)	75	6.5×10^{-11}
Possible (evens)	60	1.3×10^{-10}
Negative terms		
Possible (evens) not	40	3.0×10^{-10}
Very well possible (plausible) not	25	6.0×10^{-10}
Not likely/unlikely	15	1.1×10^{-9}
Highly (very) unlikely	2	9.8×10^{-9}
Likelihood bordering certainty not	0.1	2.0×10^{-7}

with prior or background information to give posterior probabilities for particular issues.

The following events can be defined from the previous example:

- *I* Some background information has been collected before the forensic examination. For example, data from police investigation, eyewitness statements or data from the criminal historical record of the suspect will contribute to *I*. Typically this information will reduce the number of potential suspects or objects that could be at the origin of the mark.
- *E* The evidence:features agreement (without significant differences) has been reported between the print left at a scene and a person or object under examination.
- *ID* The mark has been produced by this person or object in examination.
- $\overline{ID}$ The mark has not been produced by this object or person, and another unknown person or object is at the origin of the mark.

The definition of the two exclusive hypotheses (*ID* and $\overline{ID}$) requires consideration of the context of the case (the defense's strategy, for example); they are not always as straightforward or exhaustive as could be deduced from our example.

The Bayes' formula (Equation 3) shows how prior odds on *ID* are modified by the evidence *E* to obtain posterior odds on the issue:

$$O(ID|I,E) = \frac{\Pr(E|ID,I)}{\Pr(E|\overline{ID},I)} \cdot O(ID|I) \quad \text{(Equation 3)}$$

Where: $O(ID|I)$ = the odds on the identification before forensic examination, the prior odds on *ID* equal to the ratio $\Pr(ID|I)/\Pr(\overline{ID}|I)$;

$O(\overline{ID}|IE,I)$ = the odds on the identification given the evidence *E*, the posterior odds on *ID* equal to the ratio $\Pr(ID|E,I)/\Pr(\overline{ID}|E,I)$;

$\Pr(E|ID,I)$ = the probability of the shared features between the mark and the control, given that the mark has been left by the suspect or object in examination (*ID* is true) – generally, this value is close to 1;

$\Pr(E|\overline{ID},I)$ = the probability of the shared features between the mark and the control, given that the mark has not been left by this person or object ($\overline{ID}$ is true).

Equation 1 can also be derived in a bayesian framework, where the posterior odds on the identification considering the evidence *E*, $O(ID|E,I)$, are obtained by multiplying the prior odds of $1/N-1$ with a likelihood ratio of $1/f$.

If we agree that, in a given case, it is too conservative, or even misleading, to state that the number of potential objects or persons is maximal, then each case bears its own specificity in this regard. Hence, the estimation of the number of potential objects or persons will be based on the background information of the case (*I*) gathered by the police (investigation, eyewitness statements or data from the criminal historical record of the suspect). But generally these data are not known to the forensic examiner and are outside his or her province. Hence, the assessment of the relevant population remains the confines of the courtroom (judges, members of the jury, etc.). Consequently, before the presentation of the expert evidence, the judge will assess the size of the relevant population that could be involved in the case; this will lead to the so-called 'prior odds' on the identification. In the bayesian perspective, the open set framework covers the situation where the prior odds are estimated in reference to the total population possible (*I* represents no prior knowledge at all); any other situation with an estimate of the prior odds can be viewed as a closed set framework.

The statement (numerical or subjective) made to the court by the forensic scientist should only be the expression of the reduction factor to be applied to the prior odds (opinion of the court without the knowledge derived from the scientific evidence). Without the information about the prior odds, it is not possible for the scientist to state the fact itself (the probability that this particular person or object has produced this mark): he can only state the degree of support given to this hypothesis (or version) versus the alternative. In the bayesian framework, the strength of the evidence is given by the probability of observing the evidence under two chosen propositions. This ratio (called the 'likelihood ratio') has a numerator close to 1 and a denominator equal to the frequency of the shared features in the relevant population. The statement given by the scientist is not alone sufficient to decide on the issue (the identification or probable identification). The scientist's statement is then completely different from assessing the fact itself. The concept of evidence is therefore relative: it shows how observations should be interpreted as evidence for *ID vis-à-vis* $\overline{ID}$, but it makes no mention of how those observations should be interpreted as evidence in relation to *ID* alone. Therefore, the same piece of evidence may be insufficient proof for the court in one context, but may be the factor essential in clinching the case in another.

The odds in favour of the identification itself, that the person or object in examination has produced the mark given the circumstances of the case (background *I*) and the observations made by the forensic scientist (the evidence *E*), is a judgement based on posterior odds combined with the evidence. The essence of the decision on the identification remains the province of the factfinder.

This logical tool permits an understanding of how events interact mathematically. It clarifies the position of the scientist as well as that of the judge, and defines their relationship. The scientist is concerned solely with the likelihood ratio, whereas judges deal with the odds on the identification (**Fig. 4**). This impossibility of assessing (even by degrees) the issue in itself has already been identified in some forensic identification fields, such as handwriting examination, voice identification and tool marks.

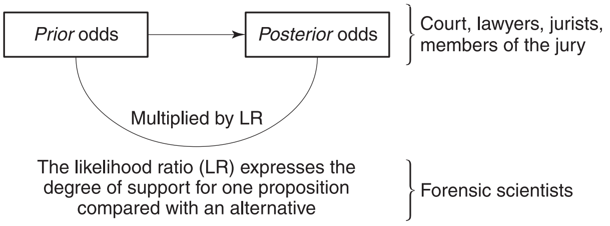


Figure 4 The Bayesian framework positioning the actors of the judicial system.

This solves the philosophical dilemma encountered by scientists asked to express a view on the identification process. Identification is obtained through a logical process that requires a judgment at the end. Such a conclusion can be drawn only by combining prior knowledge and forensic evidence. In a strictly bayesian point of view, the examiner is never in position to identify definitively. But, it may be (for fingerprints, footwear marks or DNA evidence, for example) that the likelihood ratio in favor of the identification is so strong that identification is declared, whatever prior odds might be. In such instances, a conclusion of individualization is made.

The bayesian model helps us to understand the respective duties of the court and of the forensic scientists (Fig. 4). Our schematic representation of the identification process can then be modified, taking into account the bayesian perspective (**Fig. 5**).

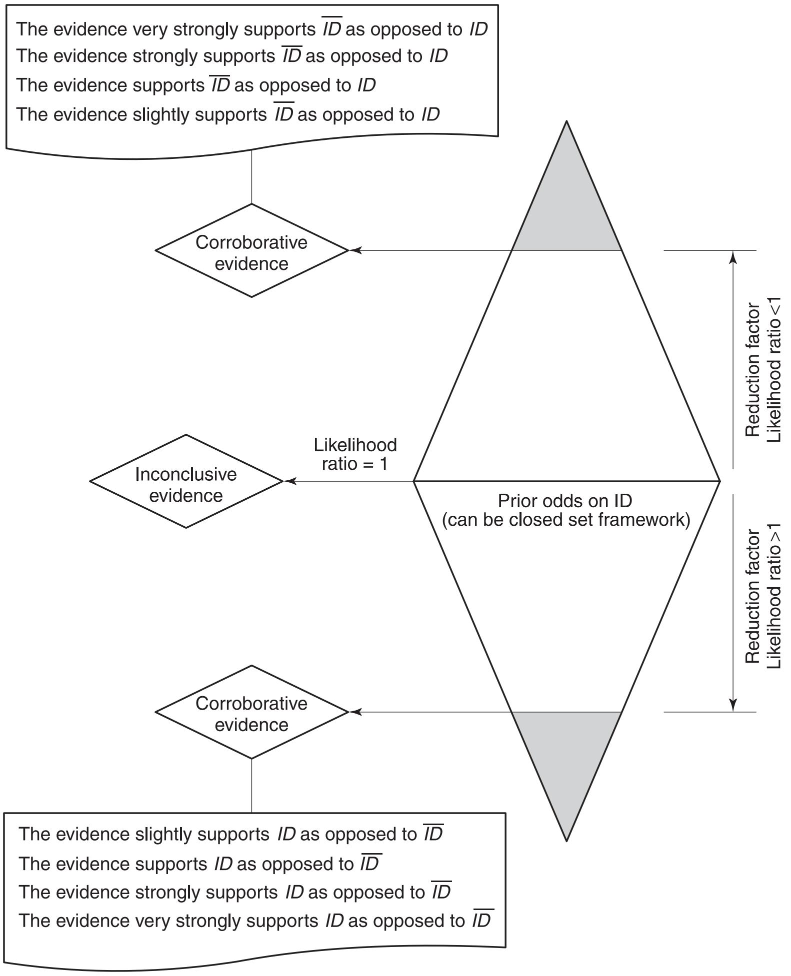


Figure 5 The value of evidence in the identification (ID) process.

Conclusion

Various decision schemes for the question of identification of sources have been presented. Two of them lead to decisions of positive identification or of qualified opinions (possible, probable, very probable identification, etc.). Both of these schemes are defined with reference to a relevant population size set to its maximum in an open set framework. The use of an open set framework forces the scientist to deal – sometimes without being aware – with thresholds to take decisions and with prior or posterior probabilities on the issue. It appears legitimate to question if these practices (or accepted assumptions) are reasonable for each particular case. Moreover, the statistical relationship between verbal opinions and frequencies of occurrence of the shared features has been shown, with all its counterintuitive and paradoxical consequences.

Another scheme, the bayesian interpretation framework, overcomes most of these difficulties, in particular by relaxing the necessity of adopting an open set framework. It provides a coherent way of assessing and presenting identification evidence. From a logical point of view, the concept of evidence is essentially relative to the case and its value is best expressed using a likelihood ratio. The question of the size of the relevant population (open set versus a closed set) depends on *prior* probabilities, which are outside the province of the scientists but belong to the court.

See also: **Criminalistics**. **Evidence**: Classification; Statistical Interpretation of Evidence/Bayesian Analysis.

Further Reading

Aitken CGG (1995) *Statistics and the Evaluation of Evidence for Forensic Scientists*. Chichester: Wiley.

Dawid AP (1994) The island problem: coherent use of identification evidence. In: Freeman PR and Smith AFM (eds) *Aspect of Uncertainty: A Tribute to D.V. Lindley*, pp. 159–170. Chichester: Wiley

Kirk PL (1963) The ontogeny of criminalistics. *Journal of Criminal Law, Criminology and Police Science* **54**:235–238.

Kwan QY (1977) *Inference of Identity of Source*. PhD thesis, Berkeley: University of California.

Robertson B and Vignaux GA (1995) *Interpreting Evidence: Evaluating Forensic Science in the Courtroom*. Chichester: Wiley.

Stoney DA (1991) What made us ever think we could individualize using statistics?. *Journal of the Forensic Science Society* **31**(**2**):197–199.

Tuthill H (1994) *Individualization: Principles and Procedures in Criminalistics*. Salem: Lightning Powder Co.

Identity Parade *see* **Facial Identification**: Lineup, Mugshot Search and Composite.

Injury *see* **Anthropology**: Bone Pathology and Ante-mortem Trauma in Forensic Cases; Skeletal Trauma.

Insects *see* **Entomology**.

Instrumental Analysis *see* **Analytical Techniques**: Separation Techniques.

International Perspective *see* **Administration of Forensic Science**: An International Perspective. **Education**: An International Perspective.

INTERNET CRIME

R L Dunne, H M Long and **E Casey**, Yale University, New Haven, CT, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0741

Introduction

This article focuses on criminal activity on the Internet in the highly intangible world of 'cyberspace'. An understanding of the Internet and cyberspace is crucial to the forensic scientist working on Internet crime, as many of the problems associated with the preservation and use of digital evidence arise from the peculiarities of cyberspace itself. The term 'Internet' as commonly used today is defined, and the relationship between 'cyberspace' and the physical world is examined. Since we are exploring new territory, 'Internet crime' itself is defined and explained. An attempt is made to identify and classify various types of Internet crimes and to assess their frequency and seriousness. The frequent interaction between Internet and real-world crime is discussed. Finally, the task of preserving digital evidence is examined.

The Internet and Cyberspace

The term 'Internet' once referred to a particular distributed group of networked computers. Increasingly, however, it has come in common parlance to be used to mean the global network of computers, the network of networks which provides its users with worldwide connectivity and communication.

The Internet, understood properly, is not a venue. It is a physical networking technology which facilitates human interaction of all sorts in a place that has come to be called 'cyberspace', a term coined by William Gibson, in his 1984 science fiction novel *Neuromancer*, to describe the virtual space in which computer-based activity occurred. Internet crime occurs *on* the Internet, but *in* cyberspace. The borderless nature of cyberspace and the difficulty of defining this new place's location, or establishing its relationship to other physical locations, is one of the seminal problems in regulating criminal activity on the Internet.

The Internet is like an enormous spider web comprising thousands of smaller webs, permitting a continuous line to be drawn between any two points on any of the smaller webs. The 'World Wide Web', the relatively recent, graphically oriented form of publication which seems destined to become the dominant vehicle for Internet interaction, is aptly named.

The development, or perhaps colonization, of cyberspace has been likened to the development of the American frontier in the latter half of the nineteenth century. It is a strong analogy. In the early days of the Internet, cyberspace was populated, although quite sparsely, by technological pioneers, feeling their way through unknown territory. They tended to be strongly individualistic and, in general, disliked and distrusted rules. They relied instead for behavioral control on a common understanding of what was appropriate and acceptable, made possible because most of these early inhabitants shared a common intellectual framework and sense of purpose. Soon, however, as the potential of cyberspace as a tool for communication and commerce began to be better understood and its population skyrocketed, the need for a more sophisticated form of control, a system of law, became apparent. It is worth noting that, although the early citizens of cyberspace are an increasingly tiny percentage of the population of cyberspace, their attitude was the seed of what has come to be called the 'hacker ethic', a belief that information should be free, which is at the root of a substantial amount of today's Internet crime.

What is Internet Crime?

There is no commonly accepted definition of 'Internet crime', although clearly not all crimes involving the use of a computer are Internet crime. The financial records of a money-laundering conspiracy might be kept, for example, on a computer. The digital evidence contained on that computer would, of course, be essential to the successful prosecution of the crime, but this would not be an Internet crime.

For the purposes of this article, Internet crime is any criminal activity that either occurs on, or makes use of, the Internet. It is important to note that Internet crime is thus not restricted to crimes committed in cyberspace and whose effects are felt only in cyberspace. Often, criminal activity on the Internet is specifically intended to produce effects in the physical world, with the Internet serving merely as a tool to facilitate the crime and avoid detection and capture. For instance, cyberstalkers use the Internet in addition to more traditional forms of stalking and harassment,

such as telephoning the victim and going to the victim's home.

Most Internet crime is the cyberspace analog of a real-world crime. Laws against fraud, for example, apply to the same conduct in cyberspace as in the real world, and proof requires establishment of the same legal elements defined by the relevant statute. Other Internet crimes are unique to cyberspace. 'Denial of service' attacks on a particular Internet site are one example. Another is breaking into a computer across the network by such means as cracking passwords, which required specialized 'unauthorized access' legislation that is now common in most jurisdictions.

A tremendously complicating factor in defining Internet crime, quantifying it and assessing its impact is the international nature of the Internet. There is no one set of applicable laws in cyberspace. When one speaks of Internet crime it matters greatly whose definition of 'crime' is used. Many Internet activities, such as gambling, that are perfectly legal in some jurisdictions are outlawed in others. Yet all jurisdictions are subsumed by cyberspace. Identifying 'all' Internet crimes by type is thus impossible, except as pertaining to a specific legal jurisdiction. Here, as we attempt to examine Internet crime from a global perspective, we must focus on the activities most commonly accepted as illegal, or at least unacceptable and problematic, by most of the Internet community.

The following activities are largely accepted as criminal misuse of the Internet:

- theft of proprietary information;
- unauthorized interception of electronic communications;
- unauthorized access to a computer system;
- theft of computer services;
- denial of service;
- distribution of a computer virus;
- financial fraud;
- espionage;
- extortion;
- trafficking in child pornography;
- solicitation of minors;
- stalking;
- harassment or threats.

Some of these crimes, such as trafficking in child pornography, solicitation of minors and stalking, are basically traditional crimes committed with the help of the Internet. For instance, the Internet enables child pornographers rapidly to transmit sexually explicit images around the world, giving commercial pornographers access to a larger market and giving cottage collectors access to a global village in which to trade. The Internet also gives sexual predators access to large numbers of children for extended periods of time, giving them ample opportunity to gain their victims' trust and arrange a meeting in the physical world. Sexual predators often take advantage of the anonymity that the Internet provides and sometimes pose as children to gain their victims' trust.

It should not be assumed that Internet crimes are more benign than their traditional equivalents. Victims of Internet crime experience the same distress as victims of crimes in the physical world. Also, some Internet crimes, such as stalking and solicitation of a minor, can lead to physical world confrontations. Although it is difficult to assess the full scope of these problems, it is clear that they have serious consequences in the physical world and are increasing in frequency as the Internet becomes more widely used.

Cyberstalking

The term 'cyberstalking' refers to stalking that involves the Internet – working in much the same way as stalking in the physical world. In fact, many offenders combine their online activities with more traditional forms of stalking and harassment, such as telephoning the victim and going to the victim's home. Cyberstalkers harass their victims using a wide variety of Internet services, including e-mail, newsgroups, chat rooms and instant messages. As well as harassing victims first encountered in the physical world, cyberstalkers target individuals they have never met. Other cyberstalkers take a less direct approach to harassment, putting personal information about their victims on the Internet, encouraging others to contact the victims or even harm them.

In general, stalkers want to exert power over their victims, primarily through fear. The crux of a stalker's power is information about and knowledge of the victim. A stalker's ability to frighten and control a victim increases with the amount of information that he or she can gather about the victim. Stalkers use information like telephone numbers, addresses and personal preferences to impinge upon their victims' lives.

Since they depend heavily on information, cyberspace is an ideal environment for stalkers, giving them access to a great deal of information about a large pool of potential targets. Additionally, a cyberstalker can determine when their victims enter cyberspace and can monitor them surreptitiously. This ability to lurk in cyberspace and protect their identity in other ways makes the Internet an even more attractive place for stalkers.

Internet computer crime

Many countries have adopted the view that certain Internet crimes are sufficiently different from tradi-

tional crime to warrant discrete legislation. This subset of Internet crime might be labeled 'Internet computer crime' because the targets of such crimes tend largely to be computers themselves and the information they contain. **Table 1** describes the most common Internet computer crimes.

Past studies have shown that a significant percentage of computer crimes were committed by individuals inside the organization whose systems were being attacked. However, the number of external attacks has grown, becoming as numerous as internal attacks. Specifically, organizations are finding that the attacks on their systems are perpetuated through their connections to the Internet. Also, there has been a dramatic increase in the number of computer security breaches overall and a corresponding increase in financial losses. As a result, Internet computer crime has become such a problem that it is considered to be a national security risk by many countries. Note, however, that Internet computer crime is still a relatively minor cause of both data and financial loss (**Table 2**).

Several attempts have been made to categorize the subset of criminals who use the Internet to commit computer crime. For instance, Icove divides computer criminals into three categories: computer crackers, computer criminals and vandals. Another study proposed the categories described in **Table 3**. However, criminals often fall into multiple categories and are perpetually finding new ways to make use of the Internet, defying the boundaries imposed by these categories. As a result, this type of categorization can add confusion to an already complex subject.

Computer cracking

Individuals who break into computer systems with malicious intent are referred to as 'computer crackers'. Crackers gain unauthorized access to computer systems in a number of ways, as described in **Table 4**. Although it takes a certain degree of skill to find new ways to implement these attacks, once a new method of attack is developed, it is often made available on the Internet to enable individuals with a minimal amount of skill to implement the attack.

Computer cracking is viewed by some as a victimless crime. However, whether a computer cracker steals proprietary information from an organization, misuses a computer system, or deletes the contents of an individual's computer, people are affected in a very real way. If, for example, a computer cracker changes prescription information in a pharmacy database, tampers with critical systems at an airport, disables an emergency telephone service or damages other critical systems, the ramifications can be fatal.

Investigating Internet Crime

To investigate Internet crime effectively it is necessary to be ever cognizant of its dichotomy. As mentioned earlier, Internet crime occurs *on* the Internet but *in* cyberspace – the components, separated by physical space, that comprise the Internet are joined to create a seamless virtual space. Therefore, when investigating Internet crime it is necessary to collect evidence stored on and transmitted using computers and, at the same time, use that discrete evidence to reconstruct the crime as it occurred in its native, seamless environment. Without an understanding of the physical components that comprise the Internet, forensic scientists will have great difficulty acquiring and analyzing evidence. Without a solid understanding of cyberspace, forensic scientists will have great difficulty assessing the significance of evidence, reconstructing Internet crimes and understanding the criminal act as a whole.

Although this is not the place for a full discussion of how the Internet functions, a summary description is warranted, starting with some basic terminology. Computers that are connected to the Internet, generally referred to as hosts, communicate using a set of protocols collectively called TCP/IP (Transport Control Protocol/Internet Protocol). Remember that the Internet is comprised of many individual networks. TCP/IP is essentially the common language that enables computers on these individual, often dissimilar, networks to communicate. Computers that are connected to two or more of these networks and direct traffic between them are called routers.

Hosts that provide a service to other computers on a network are commonly called servers, and hosts that access these servers are called clients. Any host, even a personal computer in someone's home, can become a server – all that an individual has to do is install a piece of software. Some servers allow anyone to access their resources without restrictions (e.g. Web servers) while others (e.g. e-mail servers) only allow access to authorized individuals, usually requiring a user identifier and password.

Every host on the Internet is assigned a unique number, called an Internet Protocol (IP) address, to distinguish it from other hosts. Before information is sent through the Internet, it is addressed using the IP address of the destination computer, much like an envelope is addressed before it is submitted to a postal system. Routers use these IP addresses to direct information through the Internet to its destination. If the sender requires confirmation that the destination computer has received a transmission, the TCP will perform this task, resending information when necessary. Be aware that TCP performs other functions,

Table 1 Internet computer crime incident categories

Name	*Description*
1. Compromise	Replacing or modifying part of a computer system to facilitate unauthorized access or perform malicious actions
2. Covert channels	This activity usually involves smuggling data (e.g. eavesdropped information such as captured 'sniffed' packets) out of an organization by hiding the outbound data stream (either having it (1) 'masquerade' as innocuous network traffic, (2) encrypting it, (3) encapsulating it as data inside another network protocol – also known as 'tunneling', (4) and/or using 'steganogrpahy' – a branch of cryptography where data is hidden inside other data, such as a secret message inside an image file or audio stream)
3. Eavesdropping	Sniffing network traffic and observing traffic to and from terminals as well as keyboards and monitor screens. Often a compromised machine is controlled remotely via the network, the network interface being put into a 'promiscuous mode' where it is used to covertly capture packet traffic
4. Denial of service	A purely malicious attack with the purpose of disabling access or availability of a resource (computing cycles, network bandwidth, disk space, data, etc.). Often abbreviated as DoS. Denial/disruption of service attacks may occur at the lower network layers (e.g. packets) or at the application layer (e.g. 'e-mail bombs') and may be quantitative (e.g. bombarding a network with packets to degrade bandwidth) or qualitative (a targeted attack on a Web server designed to crash or otherwise disable the specific service)
5. Harassment	Also referred to as 'cyberstalking'. Repeated unwanted communication from one individual to another. Harassment may be via private e-mail, public mailing lists, Usenet newsgroups, 'bulletin boards', instantaneous messaging mechanisms, interactive 'chat' as well as other forms of audio/video/text messaging/conferencing
6. Hijacking	The 'takeover' of, or 'piggyback' on to, a legitimate and previously established network session by a malicious individual or program. Data transfers may be redirected or modified, files damaged or stolen, login accounts compromised or malicious actions executed on the target system. In most cases the user loses total control of their 'session' (e.g. login, file transfer or Web browsing) and never regains it. Hijacking attacks are rare and sophisticated
7. Intrusion	Also known as a 'break-in', this involves obtaining unauthorized access to a computer system or network. The intruder may actually login as if on a terminal or they may just gain access to a specific network service (such as a file, e-mail or Web server)
8. Malicious software	Executables in native binary form (e.g. DOS/Windows .EXE and .COM files) as well as programs and program fragments written in scripting and macro languages. Also referred to as 'malware'. Some of the most common malicious programs (malware) are:
8a. Back door	A program which opens up access (login, dialup, network) to a machine from the outside to allow an unauthorized intruder into the machine
8b. Logic bomb	A program which is designed with a 'logic' trigger for activation of the malicious code or mode. The 'logic' is a particular condition or set of conditions. Similar to a 'time bomb'
8c. Time bomb	A program which is designed with a date/time based trigger for activation of the malicious code or mode. A 'time bomb' malicious program may be a virus, trojan or a legitimate program which stops working at a predetermined date/time. See also 'logic bomb'.
8d. Trojan	A program which does something (malicious) other than what is expected. Also known as 'Trojan horse'.
8e. Worm	A program which propagates itself (without external help), often from one computer to another across a data network (e.g. via a LAN – Local Area Network or WAN – Wide Area Network). A worm is usually standalone – not attached or in a symbiotic or parasitic relationship with another program
8f. Virus	A program which replicates itself. Parasitic, it usually attaches itself to, overwrites or replaces part of another program (the 'host' program) to spread. Major virus types are: • *Boot Sector* An infection of the boot sectors of floppy diskettes and other 'bootable' media (fixed or removable) as well as the partition sectors (e.g. MBR – master boot record) and/or DOS boot sectors of hard disks. The normal bootstrap code on the disk is replaced by a malicious version during an infection. This code runs before an OS (operating system) is loaded and run. These viruses used to be the most common, but have now been replaced in frequency by 'macro viruses' • *Companion* Virus relies on the fact that the MS-DOS command line interpreter (COMMAND.COM) invokes .COM files before .EXE files with the same base filename. • *Dropper* A program used to 'drop' a virus on to a system. Often a desirable game or free utility program available for Internet download, the dropper program or installer contains a dangerous payload – a malicious program (e.g. a virus or Trojan) is also installed on the system • *Macro* Virus written in an application's high level 'macro' or scripting language. The code is actually a part of a document or data file. The most common 'macro' viruses involve infected Microsoft Office application files (Word documents, Excel spreadsheets, etc.). 'Macro viruses' are typically spread by documents which are either e-mailed or downloaded from the Web or a network file repository

Table 1 continued

Name	*Description*
	• *Program virus* Viruses which attach themselves to either system or ordinary executable programs on disk. On DOS and Windows computers the files infected usually have filename extensions of .EXE, .COM, .OVL, .DRV, .SYS, .BIN. Other lesser affected file types have extensions of .DLL or .VXD. • *Multipartite* Viruses which can operate in either 'boot sector' or 'program' virus mode (e.g. infect either a boot sector or a program on disk)
9. Piracy	Unauthorized copying and distribution of software or other copyrighted material (e.g. audio, video, graphics, etc.). There are organized and unorganized groups on the Internet involved in software piracy (of commercial software and video game ROMs) as well as the illegal duplication of other intellectual property. The Internet jargon term for this illicit material is 'warez' (from 'wares')
10. Scanning/probing	Testing a networked computer for vulnerabilities (typically vulnerable services, but also checking for vulnerable accounts and passwords) remotely via the network. Criminality varies according to the law of each country
11. Spamming	Sending unsolicited messages, usually e-mail. Usually commercial in nature (e.g. advertising or solicitations), 'spam' messages are often sent in 'bulk' to multiple e-mail addresses. Harvesting, trading and buying/selling lists of e-mail addresses is now a business on the Internet. Legislation has been enacted in the United States to make many forms of 'spamming' illegal.
12. Spoofing	Forging/synthesizing data and/or masquerading identity at several levels: • *IP spoofing* (there is also UDP and TCP level spoofing) involves 'forging' data within packets which are then transmitted over TCP/IP networks (such as the Internet). The source of the message has usually been modified so that the real origin cannot be discovered and often is pretending to have been sent from a different real source • *DNS spoofing* involves the misdirection of Domain Name System records or servers. • *Web spoofing* is a passive attack at the application level in which a malicious Web server either (1) attempts to masquerade as another Web server or (2) 'traps' the Web browser user on the site and tricks him or her into believing they have left the site via a hyperlink – when they have not left the site and all data they are browsing as well as their responses are being passed to the malicious Web site. 'Frame' spoofing is another form of this. • *'Replay Attacks'* are a form of spoofing that involves reverse engineering fields of a captured sniffed network packet, modifying parts of the packet and retransmitting it
13. Theft of service	An attack with the purpose of obtaining unauthorized access to a resource (computing cycles, network bandwidth, disk space, data, etc.). In some cases the motive behind the 'theft' is to avoid paying (for information, Internet access, telephone service, etc.); in other cases the motive is to obtain access to a resource that is restricted or denied to the perpetrator

such as breaking information into packets, and that there are other protocols in the TCP/IP family, such as the User Datagram Protocol (UDP), the Internet Control Message Protocol (ICMP) and the Address Resolution Protocol (ARP). It is also worth noting that TCP/IP enables other protocols like Simple Mail Transfer Protocol (SMTP) and Hypertext Transfer Protocol (HTTP) to transmit e-mail and Web pages, respectively.

Whether a host is used by an individual at home to browse the Web or is used by a large corporation to manage its employees' e-mail, it contains information about the Internet activities of the people who use it. Even when information is deleted and overwritten it can be recovered. Knowing where and how to look for this information is a crucial skill when investigating Internet crime. When an Internet crime just involves e-mail, an understanding of TCP/IP is useful but not essential – investigators might only require a basic understanding of e-mail to perform an effective investigation. However, most Internet crime requires investigators to be familiar with the underlying computer and networking technology. For instance, to investigate computer intrusions effectively, a solid understanding of TCP/IP and the operating system(s)

Table 2 Causes of financial and data loss

Percent	*Description*
55	Human error
20	Physical security problems (e.g. natural disasters, power problems)
10	Dishonest employees (employees who profit from their attacks)
9	Disgruntled employees
4	Viruses
1–3	Outsider attacks

Source: Computer Security Institute.

Table 3 Computer crime perpetrator categories

Class	*Description*	*Primary motive(s)*
Corporate spies	Agents acting on behalf of corporations	Technical or business information
Hackers	Usually young (teenage or in twenties) males	Challenge, status, fame
Insiders	Employees or ex-employees	Revenge, personal gain or use
Professionals	Individual or 'organized' criminals	Information for financial gain
Spies	Agents of foreign governments	Strategic or political information
Terrorists	Amateur 'hackivists' or professional government-sponsored 'information warriors'	To cause damage or fear to achieve political objectives
Vandal	Individual or group who may or may not be angry (aggrieved party) at intended target	Damage, possibly for revenge

Source: Howard JD (1997) *An analysis of security incidents on the Internet 1989–1995*. PhD thesis, Carnegie Mellon University.

involved is required. A comprehension of the technology involved will enable investigators to recognize, collect, preserve and analyze evidence related to Internet crime.

Evidence on the Internet

When the Internet is involved in a crime, evidence is often distributed on many computers, making it infeasible to collect all of the hardware, or even the entire contents of a network, as evidence. Also, bear in mind that evidence is often present on the Internet for only a split second, making it difficult to collect. Additionally, encryption software is becoming more commonplace, allowing criminals to scramble incriminating evidence using very secure encoding schemes, making it unreadable. Furthermore, unlike crime in the physical world, a criminal can be in several places on a network at any given time. Given the many challenges that evidence on the Internet presents, a solid comprehension of how the Internet functions and how the principles of forensic science can be applied to computer networks is a prerequisite for anyone who is responsible for locating evidence and collecting it in a way that will be accepted in a court of law.

Evidence classification and individualization

In common with other forms of physical evidence, evidence on the Internet can be classified and individualized. Being able to classify and individualize evidence of an Internet crime, or the tools that were used to commit an Internet crime, can be helpful in an investigation. For example, when investigating a computer intrusion, class and individuating characteristics of the tools that were used can be helpful in determining which vulnerability was exploited, linking cases, finding additional evidence and assessing the skill level of the intruder. Some technical skill is required to closely analyze digital evidence, and knowledge of computer programming is sometimes required to decompile a program and find its class and individuating characteristics.

Virus research laboratories have classification systems for certain kinds of malicious software. However, a more comprehensive body of research classifying all forms of evidence that exist on the Internet has yet to be developed. Currently, the primary means of classifying and individualizing evidence on the Internet is through direct comparison with other samples obtained from past cases or loosely organized, incomplete archives.

Table 4 Computer crime primary internet intrusion attack methods

Attack vector name	*Description*
Authentication bypass	Gaining access while avoiding standard authentication
Authentication failure	Taking advantage of authentication systems which 'fail open'
Buffer overflows	Exploiting stack memory overwriting in networked server programs
Password cracking	Brute-force, reverse-engineering and 'dictionary'-based methods used to discover account passwords
Password sniffing	Capturing account passwords via a network 'tap'
Session hijacking	Piggybacking on authorized user connections from the Internet into internal hosts and networks
Social engineering	Impersonation of authorized personnel to gain access or network passwords
Spoofing	Having a computer masquerade as a different 'trusted' computer to gain access
Trojan horses	Malicious programs such as BackOrifice can provide 'back doors' (unauthorized avenues for access) into hosts from the Internet

See also: **Computer Crime. Electronic Communication and Information.**

Further Reading

Casey E (1999) *Digital Evidence and Computer Crime: Forensic Science, Computers and the Internet*. London: Academic Press.

Cheswick WR and Bellovin SM (1994) *Firewalls and Internet Security: Repelling the Wily Hacker*. Reading, MA: Addison-Wesley.

Dunne RL (1994) Deterring unauthorized access to computers: controlling behavior in cyberspace through a contract law paradigm. *Jurimetrics: American Bar Association Journal of Law, Science, and Technology* 35(1):1–15.

Garfinkel S and Spafford G (1996) *Practical UNIX and Internet Security*, 2nd edn. Sebastopol CA: O'Reilly.

Garfinkel S and Spafford G (1997) *Web Security and Commerce*. Sebastopol, CA: O'Reilly.

Hardy IT (1994) The proper legal regime for 'cyberspace'. *University of Pittsburgh Law Review* 55(4): 993–1055.

Hollinger RC (1997) *Crime, Deviance and the Computer*. Brookfield, VT: Dartmouth.

Icove D, Seger K and VonStorch W (1995) *Computer Crime: A Crimefighter's Handbook*. Sebastopol, CA: O'Reilly.

Pipkin DL (1997) *Halting the Hacker: A Practical Guide to Computer Security*. New York: Prentice Hall.

Rosenblatt KS (1995) *High-Technology Crime: Investigating Cases Involving Computers*. San Jose, CA: KSK.

Russell D, Gangemi GT Sr (1991) *Computer Security Basics*. Sebastopol, CA: O'Reilly.

Stephenson P, (2000) *Investigating Computer-Related Crime: A Handbook for Corporate Investigators*. Boca Raton, FL: CRC Press.

Wilding E, (1997) *Computer Evidence: A Forensic Invertigations Handbook*. London: Sweet & Maxwell.

INVESTIGATIVE PSYCHOLOGY

D Canter, University of Liverpool, Liverpool, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0780

Introduction

The domain of investigative psychology covers all aspects of psychology that are relevant to the conduct of criminal or civil investigations. Its focus is on the ways in which criminal activities may be examined and understood in order for the detection of crime to be effective and legal proceedings to be appropriate. As such, investigative psychology is concerned with psychological input to the full range of issues that relate to the management, investigation and prosecution of crime.

As Canter made clear, when he first labeled and introduced the term 'investigative psychology', its constituents can be derived from consideration of the sequence of activities that constitute the investigative process, which runs from the point at which a crime is committed through to the bringing of a case to court. This makes it apparent that detectives and others involved in investigations are decision-makers. They have to identify the possibilities for action on the basis of the information they can obtain. For example, when a burglary is committed they may seek to match fingerprints found at the crime scene with those of known suspects. This is a relatively straightforward process of making inferences about the likely culprit from the information drawn from the fingerprint. The action of arresting and questioning the suspect follows from this inference.

However, in many cases the investigative process is not so straightforward. Detectives may not have such clear-cut information but, for example, suspect that the style of the burglary is typical of one of a number of people they have arrested in the past. Or, in an even more complex example, such as a murder, they may infer from the disorder at the crime scene that the offender was a burglar disturbed in the act. These inferences will either lead them on to seek other information or to select from a possible range of actions, including the arrest and charging of a likely suspect.

Investigative decision-making thus involves the identification and selection of options, such as possible suspects or possible lines of inquiry, that will lead to the eventual narrowing down of the search process.

In order to generate possibilities and select from them, detectives and other investigators must draw on some understanding of the actions of the offender(s) involved in the offence they are investigating. They must have some idea of typical ways in which offenders behave that will enable them to make sense of the information obtained. Throughout this process they must amass the appropriate evidence to identify the perpetrator and prove their case in court.

It follows that three processes are always present in any investigation that can be improved by psychological study. First, the collection and evaluation of information derived from accounts of the crime. These accounts may include photographs or other recordings derived from the crime scene. There may also be records of other transactions, such as bills paid or telephone calls made. Often there will be witnesses to the crime or there will be results of the crime available for examination. There will be transcripts of interviews or reports from various experts. Further, there will be information in police and other records that may be drawn upon to provide indications for action. Once suspects are elicited there is further potential information about them, either directly from interviews with them, or indirectly through reports from others. In addition there may be information from various experts that has to be understood and may lead to actions. The major task of a police investigation is, therefore, typically to collect, assess and utilize a great variety of sources of information that provide accounts of crime. This is a task that can benefit considerably from the scientific study of human memory processes and other psychological studies of the reliability and validity of reports and their assessment.

The second set of tasks is the making of decisions and the related actions that will move towards the arrest and conviction of the perpetrator. There is remarkably little study of exactly what decisions are made during an investigation, or how those decisions are made. Yet there is clearly a limited range of actions available to police officers, constrained by the legal system within which they operate. From many studies of human decision-making in other contexts, it is also apparent that there are likely to be many heuristic biases and other inefficiencies in the decision-making process. Awareness of these can lead to effective ways of overcoming them.

In order for decisions to be derived from the information available, inferences have to be made about the import of that information. The third set of tasks therefore derives from developing a basis for those inferences at the heart of police investigations. These inferences derive from an understanding of criminal behavior. For appropriate conclusions to be drawn from the accounts available of the crime, it is necessary to have, at least implicitly, models of how various offenders act. These models allow the accounts of crime to be processed in order to generate possibilities for action. This process of model-building and testing is, in effect, a scientific, psychological development of the informal, anecdote-based process often referred to as 'offender profiling' or 'criminal profiling'.

A simple framework for these three sets of tasks that gives rise to the field of *Investigative Psychology* is shown in **Fig. 1**. More formally then, investigative psychology is the systematic, scientific study of:

- investigative information, its retrieval, evaluation and utilization;
- police actions and decisions, their improvement and support;
- the inferences that can be made about criminal activity, its development, differentiation and prediction;

with the objective of improving criminal and civil investigations. More detailed information about each of these three strands of investigative psychology now follow.

Information Retrieval

There are many psychological questions raised in relation to the retrieval of information during a police investigation, and various studies have led to the development of procedures to improve the information collected during an investigation.

Detail

One of the most important aspects of the information obtained during an investigation is that it should have as much relevant detail as possible. Psychologists have therefore helped to develop processes, especially for police interviews, that maximize the information obtained. In doing this, the perspective is taken that there are two issues that need to be as effective as possible. One is based on the assumption that the respondent in an interview is essentially trying to remember what occurred. Therefore anything that can help the memory process should be of value.

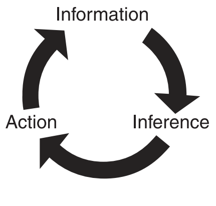


Figure 1 Investigation cycle giving rise to the field of investigative psychology.

The second issue is the relationship between the interviewer and the interviewee. If this relationship can be as supportive and helpful as possible, then more effective information is likely to be obtained.

Out of these considerations, guidelines for interviews have been developed. The best known of these is referred to as the 'cognitive interview' developed by Fisher and Geiselman. This is based on the assumption that memory is an active reconstructive process rather than a relatively passive act of recall. It draws on the well-established finding that recognition of information is much easier than its recall. Therefore any procedure that can help the interviewee to recreate the events in his or her own mind will be of value. This includes encouraging respondents to describe the events as they remember them rather than in strict response to particular questions in a given sequence. Reinstating the circumstances of the offence whenever possible, by returning to the scene or exploring details like sounds and smells, also accords with an understanding of the psychological processes by which memories are reconstructed. Attempts to consider the events from a variety of different perspectives are also considered valuable.

Investigative hypnosis has also been used to improve recall of information. In many respects hypnosis can be seen as a more intensive form of cognitive interview in which the respondent is helped to relax and concentrate. There are certainly many anecdotal accounts of its effectiveness. However, the possibilities of leading the respondent to offer information that may be suggested by the interviewing hypnotist are considered much greater than for the interviewer in a cognitive interview. Many jurisdictions therefore have very close controls over the ways in which hypnotic interviews can be conducted.

Accuracy

A number of studies have shown that the cognitive interview generates significantly more detailed information than conventional police interviews. Some studies show that the information obtained is more accurate, and also that more relevant information is obtained; but it is remarkably difficult to measure relevance or accuracy precisely, so the full value of the cognitive interview is likely to vary considerably between situations.

Attempts have also been made to use similar psychological processes to improve the recollection of faces and other details. This has proved less successful, in part because human recall of faces is so poor. Psychologists have therefore been involved in a variety of studies of how faces are reconstructed from memory and the procedures that can facilitate this. This has led to developments beyond the traditional 'photo-fit' approach. But the training involved in the use of these new systems, and their heavy reliance on effective interviewing, has meant that they have not had the uptake that would have been expected from the scientific findings.

Psychological research has also contributed considerably to the improvement in the validity of the traditional 'identity parade'. Various procedures have been introduced by many police forces around the world to ensure that the recognition task set for the witness is appropriate and not open to bias. In particular, these take account of the need to protect the suspect against the possibility of the witness's memory being modified by experiences subsequent to the crime, such as meeting the suspect in other circumstances.

Vulnerable interviewees

A number of witnesses may be regarded as vulnerable because of their age, emotional state or intellectual ability. Such witnesses may be particularly open to suggestion or may be made especially anxious or confused by the interview process. Special interview procedures have therefore been developed for interviewing such people. The procedures pay particular attention to the relationship established between the interviewee and the interviewer and the need to phrase questions and facilitate answers in ways that make sense to the respondent.

False confessions

Psychologists, in particular Gudjonsson and MacKeith, have drawn attention to the possibility that some individuals may confess to crimes they have not committed. These 'false confessions' may be a consequence of characteristics similar to those that make witnesses vulnerable, such as emotional state and intellectual ability, making the suspect more willing to accept suggestions from the interviewer. Gudjonssen has developed a measure of a person's 'suggestibility' that has been drawn on by the courts to support claims of a false confession. These may also be a product of cultural processes rather than aspects of personality, in which, for example, groups from certain ethnic minorities may deem it essential to agree with whatever a person in authority, such as a police officer, says to them. Investigative psychologists have also considered the ways in which false confessions may be produced in response to various forms of psychological or physical coercion.

Validity

In many circumstances investigators wish to assess the validity of information from witnesses because

they consider allegations may be false. If there are no objective criteria for doing this they may use one of a number of validity assessment techniques. Most of these techniques are based on the assumption that honest accounts have identifiable characteristics that are different from fabricated accounts. The most frequently used approach to statement validation is that developed by Undeutsch, known as Statement Validity Assessment, which draws upon detailed analysis of the content of a statement, referred to as Criteria Based Content Analysis. This procedure has been widely used to evaluate allegations of abuse made by children, especially in Germany, where it originated. Attempts have also been made to extend its application to statements from other groups of witnesses, with less success.

Authorship

A subset of validity questions relate to whether the words attributed to a particular author are actually the words of that person or not. This may occur, for example, when a suspect denies that he or she made the statement attributed, or in other cases of forgery or fraud. To deal with these questioned utterances, there have been a variety of attempts to use techniques based on the quantitative examination of language. These approaches are sometimes put under the general heading of 'stylistics', or forensic linguistics, or more generally forensic psycholinguistics. Yet although much is claimed for these procedures by their protagonists, the systematic research into them rarely finds any evidence to support even the mildest claims. Advances in computing techniques may change this.

These procedures are not to be confused with 'graphology', which claims to be able to provide accounts of the personality of an author from the style of his or her handwriting. There is no consistent scientific evidence for these claims.

Detecting deception

When the suspect is the source of the information, additional factors are also important beyond those of memory retrieval. These often relate to the need to determine if a person is attempting to deceive the interviewer. Thus, although there are many objective, conventional police strategies for detecting deception, most obviously determining if the known facts contradict the suspect's claims, there are a number of situations in which some knowledge of behavioral and psycholinguistic cues to deception would be very helpful. A number of researchers, most notably Ekman, have claimed that such cues are available, but others are more skeptical as to the possibility of any generally available indexes of deception from the actions or words of the suspect during a police interview.

There is much more evidence to indicate that, for many people, there are psychophysiological responses that may be indicators of false statements. The procedure for examining these responses is often referred to as a polygraph or 'lie detector'. In essence, this procedure records changes in the autonomic arousal system, i.e. emotional response. Such responses occur whenever a person perceives an emotionally significant stimulus. The most well-established indicator is when the respondent is asked to consider information that only the perpetrator would be aware of, known as the 'guilty knowledge' test. A more controversial procedure is to ask 'control questions' that many people would find emotionally significant, in order to determine if the questions elicit responses that can be distinguished from those relating directly to the crime. However, in both these applications of psychophysiological measures the most important element is the very careful interview procedure before measurements are made and during the process. In general, the technique is more productive in supporting a claim of innocence than in providing proof of guilt. For this reason many jurisdictions do not allow 'lie detector' results to be presented as evidence in court.

False allegations

In recent years there has been growing concern about the various conditions under which people will falsely allege they have suffered at the hands of others. Often, but not always, this is an allegation of sexual abuse or harassment. The various procedures for detecting deception may be relevant in these cases but, because the complainant is not a suspect, the more intrusive processes of lie detection are rarely, if ever, utilized. Instead, there have been attempts to indicate the circumstances under which such false allegations are made and use those as guidelines for more intensive examination of the circumstances; however, the validity of these procedures is still highly questionable.

Investigative Decision-Making

The main challenge to investigators is to make important decisions in often ambiguous and sometimes dangerous circumstances. The events surrounding the decisions are likely to carry a great emotional charge and there may be other political and organizational stresses that also make objective judgments very difficult. A lot of information, much of which may be of unknown reliability, needs to be amassed and digested. These are the conditions under which

various biases in investigators' thought processes are likely to occur, with consequent inadequacies in the decisions made and the subsequent actions. Recognition of the potential for these problems can lead to the development of procedures to reduce their likelihood. The challenges of police and other investigations may also be reduced by the development of decision support tools that reduce the complexity of the information that needs to be understood and assists in the derivation of appropriate inferences from the material that is available to the police.

The decision support tools that are emerging for use by police investigators each draw on particular perspectives on the nature of the problem.

Visualization

Some support tools are based on the fact that human beings can often see patterns between associations, and within activities, if they can be presented in some form of visual summary. Bar charts of frequencies are one common example of this, but commercially available software will chart networks of contacts and other sequences of associations or actions.

While these tools can be productive in summarizing a great deal of information and, in association with databases, can improve the search for and access to crucial information, they are very dependent on the skills of the particular user, often referred to in police forces as a crime analyst. In the wrong hands these systems can imply a behavioral pattern through the strong visual impact that a diagram produced, when in fact the diagram is a biased emphasis of some peripheral aspect of the criminal behavior being investigated.

Description

A further level of support to decisions can be made by identifying the salient characteristics of the offences and offenders and by producing summary accounts of them. One widespread application of this use is in the production of maps that indicate the locations where there are high frequencies of crimes, sometimes called criminal hot-spots. In these cases the salient characteristics are simply where the crimes have occurred, and the description consists of some summary or averaging of the crimes over an area in order to indicate where its geographical focus might be. All description requires some selection, distillation or averaging of information, and when that is done validly the description is helpful.

Analysis

A further level of assistance to police decision-makers can be given by carrying out some form of analysis on the crime material, typically looking for patterns of co-occurrences or discriminating nonoccurrence. An example of the former would be the recognition that certain acts of vandalism occur shortly after the end of the school day near to schools. Knowledge from descriptive analyses of the age and backgrounds of offenders prosecuted for vandalism and the geographical hot-spot information could be combined to target possible culprits and introduce other forms of crime reduction. More advanced analysis of the co-occurrence of criminal behaviors could also be used for classifying offenders and generating different investigative strategies for the different forms of offender.

Inference

When clear relationships can be established between different aspects of crimes that are of investigative interest, inferences can be made from one to the other. For example, an understanding of the relationships between where an offender offends and where he or she lives can be used to infer residential location from knowledge of offence location. The use of inference for decision support activities is at the core of the area popularly known as 'offender profiling', which will be dealt with in more detail below.

Appropriate Inferences

The traditional approach police investigators take to making inferences is the one that has always been characterized in crime fiction as 'deduction'. This is the process of reasoning from commonly known principles. For example, if a walking stick is found with strong, large teeth marks in it, then it may be reasoned that these were most likely caused by a large dog that carried the stick (as Sherlock Holmes reasons in *The Hound of the Baskervilles*). A more subtle piece of reasoning may come from the knowledge that an offender had long nails on the right hand but short ones on the left. This is a pattern favored by some guitarists, and so it may be assumed that the offender was a serious guitar player.

However, as attractive as such deductions are in fiction, they are a very poor basis for developing robust inferences in real-life crime. They are vulnerable to the knowledge and reasoning ability of the deducers and the particular features that they notice. Even more importantly, they may be worthless. It turns out that many trades give rise to people having longer nails on one hand than the other, so the inference of a guitarist could be very misleading. A dog may have bitten a walking stick in situations other than when the stick was being carried, and so

the dog may not be directly associated with the owner of the walking stick.

In order to determine what the salient aspects of an offence are, and how they may be related validly to useful investigative inferences, it is necessary to collect information across a range of cases and to test hypotheses about the actual co-occurrence of various features. This is the process of 'inductive reasoning', which is at the heart of empirical science. Investigative psychologists have consequently been active in conducting a wide range of empirical studies aimed at providing objective bases for investigative inferences. These are studies that have been characterized by Canter as attempts to solve the set of equations that link the Actions that occur during the offence, including when and where it happens and to whom, to the Characteristics of the offender, including the offender's criminal history, background and relationships to others. These have become known as the A → C equations, or the 'profiling equations'.

Studies of these equations have given rise to the identification of a number of aspects of criminal behavior that are crucial to any models of inference for effective use in investigations.

One important aspect of these models is that the variables on which they can draw are limited to those of use to police investigations. This implies that the A variables are restricted to those known prior to any suspect being identified. The C variables are limited to those on which the police can act. So, offender's personality characteristics, detailed measures of intelligence, attitudes and fantasies are all of less utility than information about where the person might be living, and his or her criminal history or domestic circumstances.

Consistency

In order to generate some form of A → C equation it is essential that the two sides of the equation are stable enough for a relationship to be established. Therefore much investigative psychology research is devoted to establishing what the salient features of an offender's crimes are, and what it is within those features that is consistent enough to form the basis of their characteristics.

It is from these studies that classification schemes are emerging considering, for example, relevant variations between serial killers and between stalkers. What is emerging from these studies is that styles of interpersonal transaction may well be consistent enough for some inferential models to be built. A distinct subset of offenders has also been identified: they have consistent relationships between their residence and where they commit their crimes, also allowing geographical inference models to be developed.

Differentiation

Although an offender's consistency is one of the starting points for empirically based models of investigative inference, in order to use these models operationally it is also necessary to have some indication of how offenders can be distinguished from each other. If every offender were consistent in the same way, then the A → C equations would provide characteristics that were the same for every offender. In part, this question reflects a debate within criminology about whether offenders are typically specialist or versatile in their patterns of offending. Research tackling this problem has tended to support the contention that the majority of chronic criminals will commit a wide range of crimes and thus cannot be considered specialist, thereby making differentiating inferences extremely difficult. However, current research is suggesting that it is possible to model offender's behavior in terms of both those aspects that they share with most other criminals and those aspects that are more characteristic of them. It is these rarer, distinguishing features that may provide a productive basis for differentiating inferences.

Development and change

A further complication to establishing the A → C equations is that the way a person commits a crime, and indeed the characteristics of a person, will change over time, even if there is a background of consistencies. However, if the basis of these changes can be understood, they can be used to enhance the inference process. In essence, the following five forms of change have been identified.

Responsiveness One important reason for differences between a criminal's actions on two different occasions may be his or her reaction to the different circumstances faced. By an understanding of these circumstances and how the offender has responded to them, some inferences about his or her interpersonal style or situational responsiveness may be made, which can have investigative implications.

Maturation This is the, essentially biological, process of change in a person's physiology with age. Knowledge of what is typical of people at certain ages, such as sexual activity, can thus be used to form a view as to the maturity of the person committing the crimes and to the basis for longer term variations in an individual's criminal activity.

Development The unfolding psychological mechanisms that come with age provide a basis for change in cognitive and emotional processes. One reflection of

this is an increase in expertise in doing a particular task. Evidence of such expertise in a crime can thus be used to help make inferences about the stages in a criminal's development that he or she has reached, and indeed to indicate the way the crimes might change in the future.

Learning Most offenders will learn from their experiences. They will therefore be expected to alter their actions in the light of the consequences of previous actions. An inferential implication of this is that it may be possible to link crimes to a common offender by understanding the logic of how behavior has changed from one offence to the next.

Careers The most general forms of change that may be expected from criminals is one that may be seen as having an analogy to a legitimate career. This would imply stages such as apprenticeship, middle management, leadership and retirement. Unfortunately, the criminology literature often uses the term 'criminal career' simply to mean the sequence of crimes a person has committed. It is also sometimes confused with the idea of a 'career criminal' – someone who makes a living entirely out of crime. As a consequence, much less is understood about the utility of the career analogy for criminals than might be expected. There are some indications that the more serious crimes are committed by people who have a history of less serious crimes and that, as a consequence, the more serious a crime, the older an offender is likely to be. But commonly held assumptions – such that serious sexual offences are presaged by less serious ones – do not have much supporting empirical evidence.

Operational applications

All of the attempts to solve the A $\rightarrow$ C equations are related to attempts to help answer the following operational questions:

1. What are the salient characteristics of the offender that will help investigators to identify and locate him or her?
2. What searches of police records or other sources of information should be carried out to help identify the offender?
3. Where, geographically, should searches for offenders be carried out?
4. Which crimes are likely to have been committed by the same offenders?
5. Which suspects are most likely to have committed the crime in question?
6. What sense can be made of the offense that will help to organize the legal case?

The complexity of deriving inferences to answer these operational questions is considerable. It requires managing issues of consistency and differentiation, together with development and change, across a range of aspects of crimes that will vary in degree of specialism. Therefore a number of investigative psychologists have collaborated with police forces around the world to develop computer-based decision support systems drawing on the ideas indicated above. These inductively developed systems are likely to replace rapidly the outmoded methods of police deduction.

The Future

Investigative psychology provides a holistic perspective on the investigation of crime, showing that all aspects of the detective's work is open to scientific psychological examination. It is helping police forces to recognize that they need to build psychological expertise into their modern computing capability, rather than just bringing an expert in when an investigation has reached a particularly difficult stage. They are learning to answer the question, 'At which point in an investigation should a psychologist be brought in?', with 'Before the crime is committed!'.

See also: **Stalking**. **Serial Killing**. **Psycholinguistics**.

Further Reading

Booth M (1997) *The Doctor, the Detective and Arthur Conan Doyle*. London: Hodder and Stoughton.

Canter D (1995) Psychology of offender profiling. In: Bull R and Carson D (eds) *Handbook of Psychology in Legal Contexts*, pp. 343–355. Chichester: Wiley.

Canter D (1995) *Criminal Shadows*. London: HarperCollins.

Canter D and Alison L (eds) (2000) *Profiling Property Crimes*. Vol. IV: *Offender Profiling Series*. Aldershot: Ashgate.

Grubin D (1995) Offender profiling. *Journal of Forensic Psychiatry* **6**(**2**):259–263.

Gudjonsson G and Haward L (1998) *Forensic Psychology*. London: Routledge.

Lee J and Alison L (1997) Beyond Cracker. *Police Review* 16 May.

IR Spectroscopy *see* **Analytical Techniques**: Spectroscopic Techniques; Spectroscopy: Basic Principles.

L

Laboratories *see* **Accreditation of Forensic Science Laboratories**. **Administration of Forensic Science**: Organization of Laboratories. **Fire Investigation**: Laboratory.

Law *see* **Legal Aspects of Forensic Science**.

LEGAL ASPECTS OF FORENSIC SCIENCE

I Freckelton, Owen Dixon Chambers, Melbourne, Australia

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0569

Introduction

Forensic science is a scientific discipline that functions within the parameters of the legal system. Its purpose is to provide guidance to those conducting criminal investigations and to supply to courts accurate information upon which they can rely in resolving criminal and civil disputes. To this extent it exists for, and also is regulated by, the law, both statutory and decisional.

Scientific analysis for forensic purposes can only lawfully take place on samples that have been properly procured. Forensic science evidence may be determined inadmissible on the basis of its provenance and, in particular, on the ground that it has been illegally or improperly obtained. If samples from the body of a suspect and a victim, be they intimate or nonintimate body samples, have been obtained by, for instance, an unlawful search or a false imprisonment of a suspect by law enforcement officers, evidence relating to the results of their scientific analysis may not be admissible. However good the scientific work, the law may prohibit the admission of the evidence by reason of the impropriety.

Scientific evidence intended to be laid against an accused person by the prosecution in criminal litigation must generally be disclosed to the accused or his or her legal representatives before trial. This is a basic requirement of fair process, so that the accused is in a position to counter it or to have a second test undertaken on the samples in order to evaluate the strength of the prosecution's case. However, scientific techniques do not always permit such a process. Sometimes, of their nature, they destroy or damage the sample available for testing, thereby precluding or impairing the potential for replication of analysis. This also has ramifications for the admissibility of the results of the testing.

Similarly, gaps in the chain of custody of exhibits later scientifically examined may lead to the inadmissibility of evidence of the results of the scientific tests because they so significantly impair the probative value of the evidence.

A number of variable factors bear upon the admissibility of the results of, and the utility of the fruits of, forensic science testing. Included among these are matters such as whether the scientist involved is sufficiently expert in the area; whether the area is one of expertise, be that determined by what is generally regarded within the area as accepted or by reference to the reliability of the theory or technique; whether the trier of fact needs or would benefit from the evidence; whether the evidence bears upon an ultimate question to be decided by the trier of fact; and whether the bases of the expert's opinion have been properly proved. For those jurisdictions where such matters are not threshold matters for admissibility, each functions as a measure by which the probative value of the forensic science evidence can be evaluated.

In many jurisdictions forensic science evidence may also be adduced without the scientist involved actually giving oral evidence, unless the person is required to be available for cross-examination by the side to whose detriment the evidence is given. This is a means of shortening proceedings and facilitating the provision of evidence where the processes of cross-examination do not need to be invoked.

Search and Seizure

Rules on evidentiary admissibility vary from jurisdiction to jurisdiction and legal system to legal system. However, it is common for courts to decline to receive evidence presented in a criminal case on behalf of the prosecution if it has been improperly obtained by law enforcement officers. By statute, and in some jurisdictions by common law, police are entitled to search the persons and the property in the possession of suspected persons. However, there are criteria for obtaining the right to engage in such intrusive exercises, be they obtaining blood or buccal samples for DNA analysis, taking pubic hair samples for matching with crime scene samples, seizing banknotes in the possession of a suspect or examining tattoos on the body of the suspect. Sometimes these involve obtaining warrants from a judicial officer on the basis of information placed on oath before the officer that there is good cause for granting such a warrant.

Where the legal criteria are breached, the fruits of the search and seizure can be described as those of the poisoned tree of an improper search and seizure. The result can be a determination that the direct fruits, namely what has been taken during the improper search and seizure, are inadmissible as evidence against the accused person. If this is so, the same result will generally follow for the indirect fruits, namely the results of scientific tests done on the improperly seized samples.

Chain-of-Custody Rules

It is essential from a legal point of view for continuity of custody of samples to be demonstrated by the prosecution from the time of their being obtained to the time that the samples are tested. It needs to be shown that there is an unbroken chain of custody of the physical sample from the time of seizure to the time of its analysis. If there is any gap in the chain, it means that the sample that was examined and tested may not be the sample seized at the claimed time from the potential scene of crime or association with the accused. The consequence is that another sample may have been tested or that there may have been tampering or contamination of the sample, resulting in inaccuracy or at least a finding unfairly prejudicial to the interests of the accused.

Ultimately, the question is one of relevance. If it cannot be proved that the sample tested was the sample seized, then the evidence is irrelevant: it does not rationally affect the assessment of the existence of a fact in issue in the proceedings. Approaching the issue in another way, if the continuity of evidence is not sufficient to establish an unbroken chain of custody, scientific evidence about the results of testing the sample is likely to be highly prejudicial against the accused person but lacking in substantial probative value. This too permits a ruling that the evidence is inadmissible.

Disclosure of Expert Reports

Rules relating to the extent of disclosure of scientific testing conducted by scientists for the prosecution and for the defense vary. Generally, it is incumbent upon the prosecution to disclose the results of testing, whether or not it favors the case for the prosecution. This prevents the defense from being taken by surprise during a trial and also puts the defense on notice of anomalies of scientific testing that may cast doubt on the evidence being laid on behalf of the prosecution.

Historically, accused persons were under no obligation under Anglo-American law to disclose any aspect of their case, including scientific evidence which they intended to adduce or upon which they intended to rely. However, in some jurisdictions, such as in some states in Australia, this is changing. Accused persons are obliged to disclose the substance of scientific evidence upon which they intend to rely at trial. Failure to make disclosure can result in judicial rulings that accused persons cannot adduce the undisclosed scientific evidence, or in costs orders against the accused because of the adjournments necessitated by the late notice to the prosecution. The change to the traditional right on the part of the accused not to have to make any form of answer to the prosecution case has been justified on the basis of the unfairness of the defense taking the prosecution by surprise and the socially unacceptable consequences of the prosecution being unable effectively to meet scientific evidence adduced on behalf of the accused.

Destruction by Testing of a Sample

Poor laboratory work or the utilization of certain tests may result in the destruction of a sample. When this occurs, it can preclude second testing of the sample by the defense, with the potential for unfairness. If a test

unnecessarily, or deliberately, destroys all of a sample, or if records are not kept which permit transparency of the testing process, this too can result in a ruling in some jurisdictions that the results of the scientific testing should be regarded as inadmissible. This is most likely where the destruction of the sample or the failure to retain records was deliberate or the result of egregious carelessness.

Evidentiary Admissibility Rules

Scientific evidence is expert evidence in the form of fact and in the form of opinion. That which is in the form of fact is admissible on the same basis as other lay opinion evidence. That which is opinion is admissible subject to evidentiary rules that differ from jurisdiction to jurisdiction. Most courts, however, require that a scientist giving expert opinion evidence be shown to be an expert by reason of specialist knowledge arising from sufficient experience, skill or training. Rigorous application of this rule can have repercussions for scientists of limited experience or whose primary expertise lies in an area of work other than that in respect of which their expertise is being asserted in a particular case. It is only when entitlement to express opinions as an expert is proved to the satisfaction of a trial judge that such opinions are admissible.

In a number of jurisdictions, reliability of theories and techniques is a condition precedent to the admissibility of scientific evidence. This is an approach that has been repudiated in England. In some jurisdictions, though, such as Canada, the US Federal Courts and some states in the United States, check-lists for ascertaining reliability have been developed, including factors such as: the potential rate of error; the existence and maintenance of standards; the care with which the scientific technique has been employed and whether it is susceptible to abuse; whether there are analogous relationships with other types of scientific techniques that are routinely admitted into evidence; the presence of failsafe characteristics; the expert's qualifications and stature; the existence of specialized literature; the novelty of the technique and its relationship to more established areas of scientific analysis; whether the technique has been generally accepted by experts in the field; the nature and breadth of the inference adduced; the clarity with which the technique may be explained; the extent to which basic data may be verified by the court and jury; the availability of other experts to evaluate the technique; and the probative significance of the evidence.

In Australia, an area of expertise test replicating the 'general acceptance test' within the scientific community exists in a number of states, meaning that the inquiry of the courts is as to whether the theories and techniques spawning test results are generally accepted within the relevant scientific community. In the United States the general acceptance test of *Frye* v. *US* continues to apply in many jurisdictions, while at federal level the Supreme Court decision of *Daubert* v. *Merrell Dow Pharmaceuticals Inc.* 125 L Ed (2d) 469; 113 S Ct 2786 (1993) has prescribed four indicia of reliability: (1) whether scientific evidence can be or has been tested, namely its falsifiability, refutability or testability; (2) whether the theory or technique has been subjected to peer review and publication as a means of increasing the likelihood that substantive flaws in methodology will be detected; (3) the known or potential rate of error and the existence and maintenance of standards controlling the technique's operation; and (4) whether a technique has gained general acceptance within the scientific community. The test therefore evaluates reliability of opinion evidence by a combination of scientific analysis and deference to the views of its legitimacy within the relevant intellectual marketplace. It has been accepted by decisions of the High Court in New Zealand.

In a number of jurisdictions, scientific evidence is not admissible if it is a matter of common knowledge, which has been variously interpreted to mean that it is inadmissible if it either does not assist the trier of fact or if it is not needed by the trier of fact. This is a function of the waxing and waning levels of understanding of scientific, statistical and other concepts by members of the general community.

In some jurisdictions, scientific opinion evidence based upon data which are not proved to the court is inadmissible. In others it is simply accounted as of minimal probative value. The reason for such an approach is that if scientists rely significantly upon others' opinions or work, the trier of fact is ill positioned to evaluate the strength of their opinions; their views are derivative and not susceptible of informed analysis by a judge or jury.

In some, albeit a reducing number of, jurisdictions, scientists and other expert witnesses are precluded from offering views on the ultimate issues to be determined by the trier of fact. This is a lingering manifestation of a concern to stop expert witnesses from 'usurping' the role of the trier of fact. It springs from a pessimistic view of jurors' capacities for reasoned evaluation of expert evidence, not overborne by the impressiveness and articulateness of expert witnesses. The rule is seldom applied in those jurisdictions where it still exists and is regarded by many commentators as ill-conceived and unhelpful.

In many jurisdictions, scientific opinion evidence can be excluded by the trial judge if admitting it

would result in evidence more prejudicial than it is probative going before the trier of fact. Criteria for applying this exclusionary mechanism differ from jurisdiction to jurisdiction, and from case to case, but have the potential to include factors relevant to the application of each of the exclusionary rules previously described. If the probative value of scientific evidence is low but its prejudicial value high, it may well be excluded as evidence. The situation becomes more complex as its probative value rises, and its prejudicial value decreases. In some jurisdictions, such as Australia and England, the tendency is for most evidence, save that which is highly prejudicial, to be admitted and the assessment of its weight is left for the trier of fact to assess.

Each of the exclusionary mechanisms is applied with greatest rigor where it is a lay jury that is the trier of fact. In those jurisdictions where a professional magistrate or judge is the trier of fact in criminal offences, without a jury, the expert evidence rules of exclusion tend to be less stringently applied.

Certification Procedures

In a number of jurisdictions, the law allows for scientific evidence to be tendered by the prosecution upon production of a report to the court and the accused person a suitable time before the trial. Often proof of the author's credentials to be accounted an expert is required. If the accused wishes to challenge the opinions expressed by the scientist, or the scientist's observations or the methodology of his or her testing, the onus is upon the accused to call for the presence of the scientist for cross-examination. Such a procedure is designed to streamline evidence that is frequently noncontentious and to save time and money. It is a procedure likely to be utilized more widely.

See also: **Expert Witness**: Qualifications and Testimony.

Further Reading

Chayko GM, Gulliver ED and Macdougall DV (1991) *Forensic Evidence in Canada*. Ontario: Canada Law Book.
Coady CAJ (1992) *Testimony: A Philosophical Study*. Oxford: Clarendon Press.
Freckelton I (1987) *The Trial of the Expert*. Melbourne: Oxford University Press.
Freckelton I and Selby H (1999) *The Law of Expert Evidence*. Sydney: LBC Information Services.
Hodgkinson T (1990) *Expert Evidence: Law and Practice*. London: Sweet and Maxwell.
Imwinkelried EJ (1997) *The Methods of Attacking Scientific Evidence*, 3rd edn. Charlottesville: Michie.
Jasanoff S (1995) *Science at the Bar*. Cambridge: Harvard University Press.
Jones CAG (1994) *Expert Witnesses*. Oxford: Clarendon Press.
Robertson B and Vignaux GA (1995) *Interpreting Evidence: Evaluating Forensic Science in the Courtroom*. Chichester: Wiley.
Stone J and Wells WAN (1991) *Evidence: Its History and Policies*. Sydney: Butterworth.
Tapper C (1997) *Cross on Evidence*. London: Butterworth.
Thayer JB (1969) *A Preliminary Treatise on Evidence at the Common Law*. New York: Kelley.

LIE DETECTION

Polygraph

F Horvath, Michigan State University, East Lansing, MI, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0747

Introduction

The use of physiological recording instruments to detect 'lying' by criminal suspects has been practiced since at least the early 1900s. Today this field is popularly known as 'lie detection' and more formally as polygraphy, the detection of deception, or, more recently, forensic psychophysiology.

In spite of its long history and widespread use, polygraphy is one of the most misunderstood and most controversial techniques in the forensic sciences. In this article, a brief history of this field is provided with respect to polygraph instrumentation and testing procedures. The current state of the field regarding

testing techniques, the uses of polygraph testing, polygraph examiners, accuracy, legal status and trends in the field are discussed.

History

Polygraph instrumentation

Development in the field can be traced along two lines, one involving changes in instrumentation and the other changes in testing techniques. Historically, the most dramatic attempts at lie detection relied upon ordeals such as placing a hot iron on the tongue of suspects, who presumably would be protected by their innocence or burned by their guilt. (Guilty persons, whose fear of detection would decrease the activity of salivary glands, would be burned because of the absence of moisture to protect the surface of the tongue. Innocent persons would not suffer such a fate.) This approach to lie detection, and similar early attempts, did not, of course, involve the use of scientific recording instruments. It was not until 1895, when Lombroso, an Italian criminologist, and his student, Mosso, used the hydrosphygmograph and the 'scientific cradle' that objective measurement of physiological changes became associated with the detection of deception. Following Lombroso, Munsterberg, a psychologist, and others noted the effect of lying on breathing, cardiovascular activity, involuntary movements and the galvanic skin response (GSR) – changes in the electrical resistance in the skin. In 1917, Marston reported success at detecting deception with discontinuous measurements of blood pressure. In the 1920s, Larson devised an instrument for making continuous recordings of both blood pressure and breathing. In the 1930s, Keeler refined Larson's apparatus by adding a device for recording GSR. In the late 1940s, Reid incorporated into the polygraph instrument a device for detecting covert muscular movements.

Although modern polygraphs represent considerable technological improvement over earlier devices, it is true, nevertheless, that the physiological measures incorporated in earlier instruments are essentially those used today. The modern polygraph is a briefcase-sized device that, in most field settings, records changes in electrodermal activity (GSR) by means of two electrodes attached to the fingertips. A standard blood-pressure cuff, which is partially inflated during testing, is used to record changes in relative blood pressure and pulse rate. Finally, changes in breathing activity are recorded by hollow, corrugated-rubber tubes, one placed around the abdomen and one around the upper thorax, which expand and contract during inhalation and exhalation. Activity in each of these physiological systems is monitored by either electronic or mechanical means and is permanently recorded on a paper chart by a pen-and-ink system. Thus, the polygraph, merely a device that enables the simultaneous recording of a number of physiological changes, is a recording device, not really a lie detector.

It is of interest to note that, because of changes in technology, it is possible to convert the polygraph's analog signals to digital form and thus to make use of computerized instruments. Such devices are widely used today and, while they have altered the way in which physiological data are displayed (e.g. on a computer monitor instead of only on a printed sheet), and to some extent how they are analyzed, the computerized polygraph is still only a recording instrument.

Polygraph testing techniques

Although commonly referred to as a lie detector, neither the polygraph nor any other device is capable of detecting a lie. This is because there is no known physiological response that is unique to lying. Lie detection is in fact an inferential process in which lying is inferred from comparisons of physiological responses to questions that are asked during the testing. It is for this reason that polygraph techniques – that is, the ways in which testing is administered and polygraphic data are interpreted – have shown more dramatic changes than the polygraph instrument itself.

There are two major testing techniques in use today, the Relevant/Irrelevant Technique (RIT) and the Control Question Technique (CQT). A third approach, generically referred to as Concealed Information Testing (CIT), is encountered in the field in special situations. Each of these procedures is in reality a family of related approaches whose specific use is governed by the situation at hand.

The RIT was widely used by early practitioners and, though in lesser favor today, it is still the procedure of choice among some examiners. Like all accepted techniques, it consists essentially of a pretest interview and polygraphic testing. During the interview the examiner discusses with the examinee background information relative to the investigation at hand, and attempts to become familiar with the examinee's language and personal history in order to ensure that the test questions are properly worded.

Essentially, the RIT consists of the asking of a series of relevant questions interspersed among irrelevant questions. Relevant questions are those pertinent to the crime at hand (e.g. 'Did you shoot John Doe?').

Irrelevant questions are not crime-related but usually ask about issues to which the examiner and the examinee know the truth is being told (e.g. 'Are you over 18 years of age?'). The series of questions may be repeated two or more times during the testing.

In the RIT, simply stated, physiological responses that accompany the relevant questions are compared to each other and to those seen at irrelevant questions. Lying is inferred from the consistency, degree and pattern of the responses.

The CQT was developed in the late 1940s by J.E. Reid. Today there are numerous variations of the CQT, all similar in principle. The CQT, like the RIT, consists of a pretest interview and polygraph testing. During the pretest interview the exact test questions are prepared and reviewed with the examinee. The question list consists of relevant, irrelevant and control questions, although other types of questions (e.g. 'guilt complex' questions) may also be asked. The relevant and irrelevant questions are similar to those asked during RIT testing. Control questions deal with matters similar to, but of presumed lesser significance than, the offense under investigation. An example of a control question in a theft case might be: 'Did you ever steal anything?' In the CQT, there are two or more repetitions of the question list during an examination.

Simply stated, in the CQT physiological responses that are consistently greater to relevant questions than to control questions indicate lying on the relevant issues. Conversely, consistently greater physiological responses to the control than to relevant questions indicates truthfulness in the matter under investigation.

The CIT procedure is limited to situations in which specific details of a criminal offense are known to the police but are not made available to the examinee. In the CIT, questions consist of a stem question with multiple options. An example of one stem question might be: 'Do you know if John Doe was killed with a ___?' The options would include the names of various weapons, e.g. gun, knife, club, etc., including the actual one used. Assuming that the examinee had no way to know which option was the correct one in a series of such questions, a greater physiological response to the correct than to the incorrect options would suggest that the examinee recognizes critical information that only the 'guilty' person would know; thus, such an outcome would indicate 'guilt'. Typically, a series of three or more such multiple-choice stem questions are asked, usually after initial testing by either the RIT or CQT. Unfortunately, most crimes do not lend themselves to such testing because valuable details about the offense either are not available or their integrity is compromised.

Uses of Polygraph Testing

In police work there are two major uses of polygraph testing. The first is referred to as specific issue testing. Here polygraph testing is used to investigate whether or not a person is involved in the commission of a particular offense. This type of testing has been shown to be extremely valuable. In addition, it is also well established that polygraph testing often exonerates persons against whom the circumstantial evidence is quite incriminating. Thus, there are great savings of investigative time and effort due to the use of polygraphy.

The second use of polygraphy is in pre-employment screening of applicants for police work. Such testing is employed by over 60% of the large police agencies in the USA and is most commonly used to investigate issues that cannot otherwise be resolved. A number of studies support the effectiveness of polygraphy for this purpose.

Polygraph testing also plays a role in the work of many intelligence agencies. It is used to determine the suitability of applicants for employment, and periodically to screen employees regarding their conduct in dealing with, for instance, persons involved with foreign governments. Finally, polygraph testing plays a role in the investigation of suspected violations of sensitive intelligence matters.

Polygraph Examiners

There are three general categories of examiners in the USA: police examiners, those employed by nonfederal law enforcement agencies; government examiners, those who work for federal law enforcement and intelligence agencies; and private examiners, those who contract for their services. The data shown in **Table 1** reveal that polygraph examiners are overwhelmingly male (95%) and college-educated (70%). (Although there are important differences between categories of examiners in some areas, the lack of space precludes discussion of them.) The great majority of them have served an internship (87%) but only a small majority (51%) work in the field of polygraphy on a full-time basis.

It can also be seen in Table 1 that 85% of examiners make use of CQT procedures as their primary method. A much smaller percentage use the RIT and other, less familiar procedures.

It is of interest to note that, in spite of the general belief that polygraph results do not play a role in judicial proceedings, 63% of examiners have testified in court regarding polygraph outcomes (Table 1). Nevertheless, only 20% believe that all polygraph examination results should be admissible evidence.

Table 1 Personal characteristics and selected practices of polygraph examiners by category

Item	*Examiner category*							
	Police		*Private*		*Federal*		*Combined*	
	n	(%)	*n*	(%)	*n*	(%)	*n*	(%)
Sex?								
Male	488	(96)	256	(95)	149	(90)	893	(95)
Female	21	(04)	14	(05)	17	(10)	52	(05)
Education?								
No college degree	132	(40)36?	50	(27)30?	3	(02)3?	185	(30)29?
College degree	139	(28)38?	67	(25)40?	41	(25)38?	247	(27)38?
Graduate work	95	(32)26?	50	(48)30?	64	(73)59?	209	(43)33?
Internship?								
Yes	430	(85)	231	(86)	157	(96)95?	818	(87)
Work effort?								
Full time	225	(46)	123	(48)	114	(74)	462	(51)
Part time	270	(54)	133	(52)	40	(26)	443	(49)
Primary technique?								
CQT	438	(89)	150	(80)74?	131	(83)	775	(85)
RIT	19	(04)	19	(07)9?	7	(04)	45	(05)
Other	37	(07)	35	(13)17?	21	(13)	93	(10)
Ever testify in court?								
Yes	295	(59)58?	198	(75)74?	88	(54)53?	581	(63)

The percentages shown have been adjusted for missing data.

This latter finding, shown in **Table 2**, which displays examiners' views on professional requirements and practices, suggests that most examiners are quite conservative on admissibility, believing that independent review of polygraph results must precede court acceptance. This is true, by the way, even though an overwhelming proportion of examiners (90%) hold the view that CQT polygraph examinations are at least 86% accurate in field settings.

Examiners are almost evenly divided on a college degree as a requirement in the field (47% agree). However, they are very supportive of a required, supervised internship (94%), governmental licensure (87%), required continuing education (97%) and some form of national certification (58%). All of these data are shown in Table 2. It is to be noted that none of these qualifications are now uniform requirements across the USA, although a few are required in some states and in some countries outside the USA.

Training of Examiners

Because of the actual nature of lie detection, the selection, training and regulation of polygraph examiners are critical to the proper development and maintenance of the field, to say nothing, of course, of the value of these items to the quality of the testing that is carried out.

The American Polygraph Association (APA), the major professional organization, currently accredits 13 training facilities. These schools must be at least 10 weeks in length and are required to devote a specified minimum number of hours of instruction to topics such as psychology, physiology and so forth. In addition, a supervised internship may be required after completion of academic instruction. In the most rigorous schools, applicants must undergo a personal interview and pass a polygraph screening examination.

About 40 states in the USA now regulate the activity of polygraph examiners. The lack of uniform regulation, particularly with respect to licensure, training, and educational requirements, is seen by many as a serious problem in the field.

Accuracy of Polygraph Testing

Field practitioners maintain that their accuracy in the field is about 90% and that errors tend to be of the false-negative rather than the false-positive type. (A

Table 2 Examiners' views on selected requirements and practices in polygraphy

Item	*Examiner category*							
	Police		*Private*		*Federal*		*Combined*	
	n	(%)	*n*	(%)	*n*	(%)	*n*	(%)
College degree required?								
Yes	183	(37)	122	(47)	129	(79)	434	(47)
Internship required?								
Yes	459	(92)	244	(96)	159	(98)	862	(94)
National certification required?								
Yes	279	(56)	147	(57)	106	(66)	532	(58)
Licensure required?								
Yes	425	(86)	241	(94)	133	(82)	799	(87)
Continuing education required?								
Yes	491	(98)	243	(94)	161	(99)	895	(97)
Should polygraph results be admissible evidence?								
All cases	95	(19)	64	(25)	26	(16)	185	(20)
With review	130	(26)	66	(26)	57	(36)	253	(28)
Stipulation	214	(44)	94	(37)	48	(30)	356	(39)
Not admit	48	(10)	16	(06)	26	(16)	90	(10)
How accurate is the CQT in actual cases?								
86–95%	238	(49)	113	(45)	65	(41)	416	(46)
96–100%	188	(38)	116	(46)	87	(55)	391	(44)

The percentages shown have been adjusted for missing data.

false-positive error is made when an actually truthful person is found to be 'deceptive' during polygraph testing. A false-negative error occurs when a person who actually lied is reported to be 'truthful'.) A great deal of controversy surrounds these claims because there is no consensus on how to interpret the available research. Another reason is that scientists do not agree on how 'inconclusive' outcomes should be considered. Do they represent errors or do they show only that, in some situations, outcomes cannot be evaluated properly and should therefore be ignored?

Although it is possible to assess accuracy in controlled settings, such as laboratory mock-crime studies, scientists disagree on whether these results are applicable to actual field settings. There are over 20 such studies now available. In these, the average overall accuracy (excluding inconclusives) is 85%, ranging between 67% and 100%, with about equal false-positive and false-negative rates, each at about 10%. If inconclusive outcomes are counted as errors, the accuracy rate is somewhat lower. These summary statistics are shown in **Table 3**.

In the field setting it is very difficult to assess the accuracy of polygraph testing. The primary reason for this is that 'ground truth' criteria (measures that

Table 3 Average 'accuracy' of control question technique in three categories of scientific studies

Laboratory: controlled 'mock crime' studies (*n*=22)	
Average overall accuracy	= 85%
Accuracy range across studies	= 67–100%
Average error: false positive	= 12%
false negative	= 10%
Field: blind review, 'proficiency' studies (*n*=11)	
Average overall accuracy	= 88%
Accuracy range	= 64–100%
Average error: false positive	= 12%
false negative	= 06%
Field: accuracy of field decisions in 'independent criterion' studies (*n*=4)	
Average overall accuracy	= 94%
Accuracy range	= 87–100%
Average error	= Not specified

establish with certainty who is and is not actually telling the truth) are often lacking. This problem, of course, is one that must be confronted in almost all assessments of test outcomes in real life, and it is a difficult one to deal with.

Research on the accuracy of the CQT in real life has been approached in two different ways. The first of these has been referred to as blind-review or dependent criterion studies. (These are very similar in nature to what is done in proficiency studies in other forensic science research.) The earliest example of this approach with the CQT was reported in 1971. In that study examinations were carried out in a number of independent, actual cases. In each, some examinees 'passed' (that is, they were reported to be truthful) and some were found to be 'deceptive' to the issue under investigation. The deceptive person in each investigation was questioned after the polygraph testing and each made a full confession, exonerating the other suspects in the case. Thus, in each case the polygraphic data of each suspect were confirmed as being either truthful or deceptive by the confession of the 'guilty' person.

In these cases then, the examinees' truthfulness or deception had been 'confession verified'. For each, complete polygraphic examination data (records) were available. From these, the records of 20 truthful persons and 20 deceptive persons were selected and each was independently evaluated by ten examiners, none of whom was involved in the actual testing. These examiners were asked to determine which were the records of truthful persons and which of deceptive persons. Overall, the ten examiners produced an average accuracy of 88%.

More than ten generally recognized studies similar to this one have been reported in the literature. While the findings in these studies have varied somewhat, similar to laboratory-based studies, the average 'accuracy' has been 88%, as shown in Table 3. In some studies, false-positive rates have been as high as 50%, whereas false-negative rates generally were much lower; on average, these rates are 12% and 6%, respectively. The results from such research are controversial and there is no consensus on what they reveal about accuracy in real-life cases.

Another approach to estimating the accuracy of polygraph testing has involved attempts to establish ground truth by use of criteria independent of the polygraph examinations themselves. For example, in one of the most recent of these studies, researchers attempted to verify the outcome in cases in which polygraph examinations had been carried out by police. They reportedly were able to verify the polygraph examiners' outcome in 89 instances; in these, the examiners' decisions were correct 94% of the time. On actually truthful persons the examiners' decisions were correct 90% of the time; on deceptive persons, decisions were 100% correct. Similar results have been shown in several other attempts to confirm examiners' decisions in the field by use of criterion measures independent of the examination process. Across these studies the average overall accuracy, as shown in Table 3, has been 94%, ranging between 87% and 100%.

There is also considerable disagreement about which of these independent-criterion studies is of value and, as with the blind-review studies, how best to interpret these findings. One concern, as an example, has to do with how accurately the independent criterion assesses examinees' actual truthfulness or deception. Thus, in spite of the available scientific data, the controversy over the accuracy of polygraphy continues.

Relative Accuracy of Polygraphy

In spite of the lack of scientific consensus about the accuracy of polygraph testing, it is seldom that the issue is placed in proper perspective. One important question to be asked is not how accurate polygraph testing is in the abstract, but rather how accurate it is relative to other types of evidence and other processes used to accomplish similar objectives. When considered in this light, the evidence is quite different from that usually presented. CQT polygraph testing in controlled conditions, for instance, shows an accuracy that equals or exceeds that of other common means of investigation. In one study it was shown that polygraph testing produced an accuracy that was comparable to results obtained by document examiners and fingerprint analysts, and greatly exceeded that of eyewitnesses. Equally important, polygraph testing was shown to have greater utility, that is, it was useful in more situations than other forms of similar evidence considered.

A comparison of the accuracy of polygraph testing with other common forensic techniques shows similar results. For example, in recent years, in the USA at least, proficiency testing of widely accepted and extensively used crime laboratory techniques has become an accepted method of assessing performance. Generally, in this testing, crime laboratory personnel are asked to analyze in the blind, samples of physical evidence which have been specially prepared for study. Assessments of these samples by practicing crime laboratory specialists are then returned to a central site where they are matched against known standards. In this way, the proportion

of agreements and disagreements of the experts' decisions with the standards can be determined and the proficiency with which certain kinds of physical evidence are analyzed can be specified.

In one of the earliest proficiency studies, carried out in the early 1980s, the results shown in **Table 4** were reported. It can be seen in that table that the proportion of laboratories submitting 'unacceptable responses' (returns that did not match known standards) varied widely for different types of evidence, from 1.6% for examination of certain fluids to 71.2% for examination of certain blood samples. In other words, common forensic techniques were shown to have widely varying accuracy rates.

The results in more recent proficiency studies, reported in 1995, have shown some improvement in specific areas but there has not been dramatic change overall. For example, **Table 5** shows representative findings pertaining to three forensic tests, from which it can be seen that the accuracy with which certain samples of common evidence (blood stains, automobile paint and questioned documents) were correctly identified varied between 64% and 89%; false-positive rates in these tests ranged between 7% and 49%. Although not shown in Table 5, the accuracy range for identification of specific drug samples was between 47% and 100%, and between 85% and 98% for samples of blood and other body fluids.

What such studies demonstrate is that all forensic testing is fallible, that accuracy statistics vary considerably, depending on the technique evaluated, and that polygraphy, when considered in relation to other commonly used forensic techniques, yields comparable and, in some cases, superior accuracy.

As a final point, crime laboratory proficiency studies have reinforced the same lesson that has been made clear in the studies dealing with polygraphy. When tests are carried out by poorly trained examiners, when improper procedures are applied, or when other significant violations of accepted standards are present, the outcomes are less accurate than would otherwise be the case.

Table 4 Percentage of crime laboratories submitting results of 'unacceptable proficiency' in various forensic analyses: ordered by percentage of unacceptable responses

Sample number	*Sample type*	*No. of laboratories responding with data*	*No. of 'unacceptable' responses*	*Laboratories submitting 'unacceptable' responses (%)[a]*
13[b]	Physical fluids (b)	129	2	01.6
6	Drugs	181	3	01.7
12	Fibers	120	2	01.7
13[a]	Physical fluids (a)	129	3	02.3
3	Blood	158	6	03.8
4	Glass	129	6	04.8 $$$4.7
7	Firearms	132	7	05.3
20	Questioned documents (a)	74	4	05.4
1	Drugs	205	6	07.8 $$$2.9
21	Firearms	88	12	13.6
15	Drugs	143	26	18.2
	Questioned documents (b)		14	18.9 $$$100.0
5	Paint	121	24	20.5 $$$19.8
19	Wood	65	14	21.5
17	Metal	68	15	22.1
18[b]	Hair (b)	90	25	27.8
2	Firearms	124	35	28.2
14	Arson	118	34	28.8
9	Glass	112	35	31.3
16	Paint	103	35	34.0
11	Soil	93	33	35.5
18[e]	Hair (e)	90	32	35.6
18	Hair (a)	90	45	50.0
10	Paint	111	57	51.4
18[c]	Hair (c)	90	49	54.4
18[d]	Hair (d)	90	61	67.8
8	Blood	132	94	71.2

[a] $\frac{\text{Number 'unacceptable' responses}}{\text{Number Laboratories responding with data}} \times 100 = \text{Percent unacceptable.}$

Revised from Peterson J (1983).
The percentages shown have been adjusted for missing data.

Table 5 Correct and incorrect results (%) in three representative forensic tests in recent (1995) crime laboratory proficiency studies

Evidence type	*Correct*	*Wrong*	*Inconclusive*	*False negative*	*False positive*
Blood stains	89	6	5	2.2	11.3
Automobile paint	74	23	2	6.3	49.1
Questioned documents	64	3	32	2.3	7.4

Legal Status of Polygraph-Testing Results

Although polygraphy is widely used informally on a daily basis in the US justice system, the admissibility of polygraph testing results in court is in a state of flux. This is partly due to two recent decisions by the US Supreme Court, *Daubert* v. *Dow Merrell Pharmaceutical Company* and *US* v. *Scheffer*. The former decision, the *Daubert* case in 1993, overturned the 1923 *Frye* v.*US* decision, which established the 'general acceptance' rule for all scientific evidence, and gave judges greater latitude in deciding which forms of scientific evidence to admit at trial. The more recent 1998 *Scheffer* decision dealt directly with the issue of admissibility of polygraph testing results in military trials. In this case the court ruled that there was still not 'scientific consensus' regarding polygraphy, and for that reason a presidential directive prohibiting admissibility in military trials was legally permissible.

Both the majority and minority views in the *Scheffer* case raised some important issues regarding polygraphy. In the former instance, the majority hinted that, in spite of the current scientific controversy about accuracy, when the research data are clearer the court may reconsider admissibility. In the latter instance, it was noted that the government argued against admissibility largely based on a claim of unacceptable accuracy. Yet, it is the government that is one of the major consumers of polygraph examinations, even in the most highly sensitive and complex cases involving national security. This inconsistency in the government's position suggests that the issue of accuracy is not at the heart of opposition to admissibility of polygraphy. Other issues raised in this case, as well as in other cases in which admissibility has been considered, may be weighted more heavily. One of these, for example, is the belief that the admission of polygraph examination results will produce unnecessary collateral litigation; there will be simply too much court time spent on determination of admissibility in individual cases. Another is that the inclusion of such results in court intrudes on the 'trier of fact', the jury. However, several scientific studies, as well as juror polls, show that jurors are not overwhelmed by polygraph examination results. Moreover, proper court treatment of polygraph testing can ensure that it is given the weight it deserves in a particular case. This is consistent with the role of the court in dealing with other types of scientific evidence.

In spite of the general exclusion of polygraph evidence, there appears to be a growing tendency in federal and state courts for cautious acceptance. In most of these cases, admissibility is predicated on the establishment of rather stringent control over the qualifications and experience of the examiner and the process used in the administration of the examination. One reason for the growing acceptance is that polygraph examinations play an increasingly important role in the justice system. For instance, almost all major police agencies in the USA employ one or more polygraph examiners; they carry out examinations in almost every major criminal investigation and are quite successful, even in those cases in which other means of investigation are either not possible or are quite limited. Prosecutors and defense attorneys routinely make use of polygraph examinations and judges regularly use them, if not formally, then informally, to help decide issues of interest in pretrial, trial and posttrial stages. The evidence suggests that these uses will continue to grow.

Worldwide Use of Polygraphy

The USA has between 3000 and 4000 active examiners. Growth in the field is primarily outside the USA. There are 56 countries in which polygraph testing is now being used; 18 of these, including Canada, Israel, Japan, Romania and Singapore, have very active polygraph testing programs. Training schools have recently been held in Russia, South Africa, Turkey, Mexico, Venezuela and Singapore. In addition, personnel from a wide variety of countries, including China, have been trained in the USA in the past decade or so. These developments will certainly lead to more widespread acceptance and use of polygraphy.

See also: **Education**: An International Perspective. **Legal Aspects of Forensic Science**.

Further Reading

Abrams S (1989) *The Complete Polygraph Handbook*. Lexington, MA: Lexington Books.

Ben-Shakhar G and Furedy J (1990) *Theories and Applications in the Detection of Deception*. New York: Springer.

Bersh P (1969) A validation study of polygraph examiner judgments. *Journal of Applied Psychology* **53**:399–403.

Department of Defense (1984) *The Accuracy and Utility of Polygraph Testing*. Washington, DC: US Department of Defense.

Horvath F (1995) *Professionalism in polygraphy: organizational membership and its relationship to views, attitudes and practices in the field*. Paper presented at a meeting of the American Polygraph Association. Las Vegas, Nevada: American Polygraph Association.

Horvath F (1997) *Scientists and practitioners' views on scientific issues in the use of applied psychophysiological methods to determine credibility: from Frye to Daubert and into the 21st century*. Paper presented at a meeting of the American Polygraph Association. San Diego, California: American Polygraph Association.

Horvath F and Reid J (1971) The reliability of polygraph examiners' diagnosis of truth and deception. *Journal of Criminal Law, Criminology and Police Science* **62**(2):276–281.

Marston W (1938) *The Lie Detector Test*. New York: Richard R Smith.

Matte J (1996) *Forensic Psychophysiology Using the Polygraph*. Williamsville, NY: JAM Publications.

Office of Technology Assessment (OTA) (1983) *Scientific validity of polygraph testing: a research review and evaluation – a Technical Memorandum*. (OTA-TM-H-15). Washington, DC: US Government Printing Office.

Peterson J (1983) *The Proficiency of Crime Laboratories: Taking a Hard Look at Laboratory Results*. Based on results of Crime Laboratory Proficiency Study. Washington, DC: National Institute of Justice, US Department of Justice.

Peterson J and Markham P (1995) Crime laboratory proficiency testing results, 1978–1991. I. Identification and classification of evidence. *Journal of Forensic Sciences* **40**(6):994–1008.

Peterson J and Markham P (1995) Crime laboratory proficiency testing results, 1978–1991. II. Resolving questions of common origin. *Journal of Forensic Sciences* **40**(6):1009–1029.

Raskin D (ed.) (1989) Polygraph techniques for the detection of deception. In: *Psychological Methods in Criminal Investigation and Evidence*, pp. 247–296. New York: Springer.

Reid J and Inbau F (1966) *Truth and Deception: The Polygraph ('Lie Detector') Technique*. Baltimore, MD: Williams and Wilkins.

Widacki J and Horvath F (1978) An experimental investigation of the relative validity and utility of the polygraph technique and three other common methods of criminal identification. *Journal of Forensic Sciences* **23**(3):596–601.

Linguistics *see* **Psycholinguistics**.

Lip Print *see* **Cheiloscopy**.

M

Major Incidents *see* **Crime-scene Investigation and Examination**: Major Incident Scene Management.

Management *see* **Administration of Forensic Science**: An International Perspective; Organization of Laboratories.

Manslaughter *see* **Crime-scene Investigation and Examination**: Collection and Chain of Evidence; Contamination; Criminal Analysis; Fingerprints; Packaging; Preservation; Recording; Recovery of Human Remains; Scene Analysis and Reconstruction; Suspicious Deaths.

Mass Disaster *see* **Disaster Victim Identification**.

Mass Spectrometry *see* **Analytical Techniques**: Mass Spectrometry.

MICROCHEMISTRY

S Palenik, Microtrace, Elgin, IL, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0574

Introduction

Microchemistry originally referred to chemical reactions conducted on a small scale where the quantity of reactants was measured in micrograms. In the forensic sciences, it has acquired a broader meaning and today refers to microchemical reactions which are used to establish the elemental and/or chemical composition of physical evidence; either to help establish identity or provide a chemical basis for comparison. It is distinguished from instrumental analytical techniques such as Fourier transform infrared (FTIR) microspectrophotometry, energy dispersive X-ray (EDS) spectroscopy and mass spectrometry (MS), which have the same goals but utilize physical rather than chemical properties. In some areas of forensic science, such as drug analysis, instrumental methods have almost entirely replaced microchemical ones. In other applications, microchemical tests can provide information that is difficult or impossible to obtain from instrumental methods alone. For example, acrylic automotive lacquers can be readily identified from their FTIR spectra. The spectrum will not, however, distinguish between a solution and a dispersion lacquer. Microchemical solubility testing of

the paint can answer the question. Dispersion lacquers are soluble in both acetone and xylene, whereas solution lacquers dissolve in acetone but are insoluble in xylene. In some cases, microchemical reactions are so fast, certain of result, and require so little sample preparation that they are the method of choice. The identification of lead in possible bullet holes or ricochet points using sodium rhodizonate is a striking example. The reagent forms a bright red precipitate in the presence of lead.

Types of Microchemical Reactions

Because microchemistry is a broadly defined concept as it is used in forensic science, many different types of microchemical manipulations and reactions are categorized as microchemistry. Thus it includes both color reactions and microcrystal tests for drugs, as well as solubility tests on paint films, in addition to tests which are specific for a particular inorganic ion or organic functional group. This versatility is one of the strengths of classical microchemistry; it can be applied to inorganic, organic and complex substances, unlike some instrumental methods which only provide information of one kind; for example, EDS spectroscopy, which only identifies elements. One of the disadvantages of microchemical analysis is that proper interpretation of the results usually depends upon an understanding of the chemistry of the reactions. In the case of most microchemical tests for drugs, however, presumptions about the identity of the drug(s) present can be made even though the chemistry of the tests is not well understood. Thus, in a police shooting in New Jersey, an analyst stated that one of the holes in a parka shell was not from a bullet, because he failed to detect lead with a spot test. A reexamination of the hole revealed the presence of dried body fluids and fatty substances surrounding it. When these were destroyed by warming with an oxidizing acid, the residue gave a positive reaction using the same test. The identity of lead was then confirmed by EDS spectroscopy. This illustrates the principal reason why some analysts cast a dubious eye on microchemical tests. To those unfamiliar with their use, they appear unreliable. In the hands of a well-trained and experienced microanalyst, they are another powerful analytical tool, to be used along with the latest microanalytical instruments.

Microchemical reactions can be classified on the basis of their reaction products, whether they are directed to inorganic or organic analysis, and the technique used to perform them. When classified on the basis of technique, most microchemical tests can be divided into one of two methods: spot tests and crystal tests.

Spot tests

These are typically based upon the formation of colored reaction products, but reaction indicators involving the formation of gases, fluorescent compounds, etc. are also utilized. The tests are typically performed in the depressions of glazed white or black porcelain plates (spot plates) or on filter paper. The field of spot test analysis was developed by Fritz Feigl of Austria, who was concerned with specific, selective and sensitive reactions for both inorganic and organic chemical analysis on a small scale. Although so-called color tests were used before Feigl began to publish his work, they were usually of an empirical nature, with the reason for the color changes and the nature of the chemical reactions which caused them understood only poorly or not at all. Specific reactions are those in which a positive result indicates the presence of the sought-after substance (cation, anion, functional group) without any doubt. Selective reactions give the same analytical result with more than one substance, but the reaction mixture can be conditioned by the addition of other chemicals, to prevent interfering ions or compounds from reacting and thus be made virtually specific. The sensitivity of a microchemical reaction is stated in terms of its *limit of identification* and *limit of dilution*, which are the ultimate amount of the substance which can be detected by the reaction and the limit at which it can still be detected in solution, respectively. Detection limits are typically in the microgram to nanogram range and are thus suitable for detecting the major components of small particles or small amounts of unknown material when testing is conducted on the proper scale. In many cases, published detection limits can be exceeded by performing the reaction in a confined area (in a narrow bore capillary tube, on the tip of a cellulose fiber, etc.) and examining the reaction product under a microscope. Fluorescent reaction products greatly extend the sensitivity of microchemical tests, as very low intensity fluorescent light can be efficiently collected by microscope objectives of high aperture.

Spot tests are usually very sensitive, but very few specific tests exist and most of these are for inorganic ions rather than organic functional groups or compounds. One of the common uses of spot tests is as screening tests for drugs of abuse. Most of the tests employed for this purpose are not specific for a particular drug but detect a functional group which is common to one or more illegal drug compounds. In this case, the value of the test is that it provides the information necessary to decide if testing should be carried further or halted; there normally being no need to continue an analysis if scheduled substances are absent. It is necessary that spot tests used for this

purpose be reliable and that both positive and negative results should be easy to interpret, otherwise the decision to undertake or abandon further testing will be based on an unreliable assumption and mistakes made. Although it is desirable that the chemistry behind such screening tests be understood, it is not absolutely necessary if the only purpose is screening. However, when spot tests are used to conduct a chemical analysis and identify material of unknown composition, it is absolutely essential that the chemical basis for the reaction(s) be fully understood. Because spot tests can usually be applied to mixtures, without preliminary separation or purification, they can also be useful for the identification of excipients (cutting or diluting agents). Spot tests are also commonly used to locate and identify compounds on thin-layer chromatograms after they have been separated.

Beside their use as screening tests for controlled substances, spot tests are used throughout a modern forensic science laboratory as well as in the field at crime scenes. Some of these applications are described below.

Spot tests are used for explosive and explosive residue analysis. Microchemical tests on solvent extracts or individual particles handpicked from explosive debris can give useful information about the presence or absence of undetonated explosive compounds as well as explosion reaction products from both high and low explosives. Color tests are available for high explosive organic compounds, such as RDX, HMX, dinitrotoluene, nitrocellulose, nitroglycerin and trinitrotoluene (TNT), as well as low explosive components and postexplosion residues, such as aluminum, ammonium ion, antimony, bromide, chlorates, chlorides, magnesium, nitrates, nitrites, perchlorates, phosphorus and sulfur. The selectivity (and in some instances specificity), sensitivity and ease with which these tests can be performed make them useful in both the laboratory and in the field. Preliminary tests can be followed up with more specific microchemical reactions or instrumental techniques depending on the results obtained. Since many organic explosives have a significant vapor pressure, it is possible to use spot tests to detect residues from hands and clothing by drawing vapor from an enclosed area containing the body part or item (enclosure in a nonpermeable plastic bag works well) through a tube containing glass beads or ion exchange resin beads which have been treated with a reagent to detect the compound being sought.

Unburned powder residues (often referred to as 'powder burns') on clothing or skin can be visualized and identified by spot testing using the Griess reaction, which is specific for nitrites and selective for nitrates. Sheets of photographic or filter paper treated with the reagents can be used to 'map' the location of the residue, even on multicolored or dark clothing where observation is otherwise difficult or impossible. Spot tests for lead and copper are useful for identifying these metals in suspected bullet holes or at the point of a suspected ricochet. As mentioned earlier, sodium rhodizonate is a sensitive reagent for lead. Enough lead (II) ions can be brought into solution to give a positive test by treating the metal with dilute (5%) acetic acid. Weak ammonium hydroxide ($\sim$10%) can dissolve enough copper metal to react with rubeanic acid to give an olive green precipitate of copper rubeanate. The tests are specific for these cations. Since the tests can be performed *in situ*, they are easy to use at a crime scene. Such testing has proven its value in practical investigations. Microscopical and microchemical analysis of a gray residue, isolated from a bullet removed from a victim in a suspected police shooting, indicated that it was composed of weathered concrete. The site at which the shooting was alleged to have occurred was a small park, along one side of which was a brick wall. Based on statements given by possible witnesses to the predawn incident and a preliminary reconstruction of the scene, suspected ricochet marks in the mortar between the bricks were subjected to spot tests for lead. One of these marks gave a positive test for lead which provided a benchmark to help decide among several possible bullet trajectories and test the truth of the statements made by the witnesses.

The use of spot tests in serology is perhaps one of the oldest uses of such tests in forensic science. Most of the preliminary tests for blood are based on reactions in which hydroxyl ions, formed by the enzymatic action of heme, reoxidize the reduced reagents in the presence of hydrogen peroxide. The most commonly used reagents of this type are phenophthalin, leucomalachite green (formed by reducing phenolphthalein and malachite green, respectively, to their colorless leuco compounds with zinc) and derivatives of benzidine. The last-named reagent is available in the form of test strips which can be moistened and placed against a spot suspected of being blood. A rather sensitive, but still only selective, test for blood is based on the luminescence of luminol. It is used most commonly when large areas must be searched for blood or blood patterns or in locating a trail of occult blood. Photography of the chemiluminescence produced by this reagent in the presence of even minute amounts of blood can provide convincing evidence which would be otherwise difficult to document. The most minute particles of blood, such as a stain on a single hair, can be identified using

fluorescence. The specimen is placed on a microscope slide and brought into focus with transmitted light. The transmitted light is blocked off and the specimen is excited with filtered ultraviolet light from a mercury burner attached to the microscope. A drop of sulfuric acid is allowed to flow between the slide and coverslip. If blood is present, the acid dissolves the iron from the hemoglobin. In the absence of iron, the empty porphyrin molecule exhibits a bright red fluorescence which can be detected from the smallest particles of dried blood. Spot test reactions for other body fluids such as urine, semen, sweat and saliva have also been developed and are used when necessary to detect and identify these substances.

Spot tests find their principal forensic application in the analytical examination of trace evidence. In addition to well-recognized items, such as fibers, paint, glass, soil, etc., materials of unknown composition are frequently encountered in the course of evidence search and recovery operations or are simply submitted as unknowns. In the examination of fibers, for example, spot tests can be used to determine the degree of lignification of vegetable fibers as an aid to their identification, permit the distinction between cotton fibers which have or have not been resin treated, or distinguish between regenerated fibers of lyocell or viscose by detection of the carbon disulfide residue from the viscose process. The reaction of nitrocellulose with diphenylamine and sulfuric acid, while barely selective, let alone specific (it reacts with almost all oxidizing agents), can be regarded as specific when applied to a paint flake or smear. Color reactions are easily applied to aqueous extracts to determine the presence and relative amounts of fertilizer compounds as well as the pH of small soil samples.

The intelligent application of microchemical tests, especially when used in conjunction with microscopy and microanalytical instrumentation such as micro-FTIR, EDS and MS, can provide chemical information and lead to an identification of almost any unknown material that may be encountered. A systematic procedure based on observation and experiment and not a flow chart approach will rarely fail to provide at least some information about substances of unknown composition. The microanalyst should never perform a test unless he or she has a clearly formed question which the test is expected to answer. The next step in the analysis is decided by the results, positive, negative or inconclusive, of the first test or observation after it has been interpreted. In this way the instrument, technique or test most likely to yield the most useful information can be selected, while conserving what is often a minute and irreplaceable specimen.

Crystal tests

Crystal tests derive their name from the fact that the reaction products of these tests are crystalline solids. A positive reaction results in the precipitation of a characteristic crystal form or habit which can be recognized under the microscope. In some cases characteristic optical properties, such as color (including pleochroism), anomalous polarization colors, extinction angle or an unusually high or low refractive index or indices, may also be used to distinguish between otherwise similar crystals formed with ions or compounds other than the one being sought, or to confirm an identification. In most cases, however, it is the morphology of the crystals that is used to recognize a positive test. Crystal tests have been developed for both inorganic cations and anions as well as organic compounds.

Qualitative crystal tests for inorganic ions are well established and an extensive literature exists documenting the technique, chemistry and reliability of these reactions. Because only ions with a particular size (ionic radius) and charge can fit in a particular crystal lattice, these tests tend to be specific or at least more selective than corresponding spot tests, although, owing to solubility considerations, they are normally less sensitive. Crystal tests are typically conducted on microscope slides. Various techniques are employed for conducting the reactions. The common method is to dissolve 'period-sized' particles of the test substance and reagent in small droplets of water on a slide. The drops are then joined together using a thin glass rod to connect them. No coverslip is used. Where the drops first meet, the concentration of both reactants is highest. Crystallization here tends to be rapid and the crystals may be small and poorly formed. As the reactants diffuse through the test drop, their concentrations decrease with respect to each other until well-formed and slowly grown crystals begin to precipitate somewhere on the slide. Crystals usually form rapidly in inorganic tests, tend to be reasonably large and can be located and identified using total magnifications of only about $100\times$. Because these tests are based on well-known and understood chemical reactions, they are certain of result, cost little to perform and can be rapidly performed. The chemical information they provide is best used to confirm an identification and not as part of systematic analyses. The latter can usually be performed more efficiently using instrumental techniques, although they are the method of choice for the qualitative identification of acid radicals and certain other difficult to identify chemical species, such as the ammonium ion. Thus they can be used to rapidly identify a metal particle after dissolving a portion of it

in a droplet of acid, confirm an identification made by optical crystallography with the polarizing microscope or used to test for the presence or absence of a specific substance, usually without purification or separation. Because of the simple apparatus required, crystal tests lend themselves particularly well to the needs of the forensic microscopist. Microscopic specimens can be removed from mounting media, and microchemical reactions can be performed as the analyst watches through a stereomicroscope.

In some instances, the general class to which the substance belongs is known or suspected, for example drugs or paint. Crystal tests have been used for identification of organic drug compounds since the second half of the nineteenth century, primarily through the work of Wormley in America and Helwig in Europe. Most of these tests are based on the formation of addition compounds between reagents consisting of heavy metal ions (e.g. platinum, gold, etc.) and halides (iodides, bromides and chlorides) with the large nitrogen-containing drug molecules. One advantage of these tests is that they can be applied to relatively impure specimens because the reagent will only seek out, react with, and precipitate the nitrogen bases that are present in the mixture. Another is that they can be designed so that they can distinguish optical isomers. The introduction of instrumental methods of analysis after World War II saw a gradual decrease in the use of microcrystal tests, as they came to be called. Gradually, as scientists trained in their use and interpretation became scarce, a general distrust of the technique arose, primarily from chemists whose training in analysis was entirely based on instrumental methods. In fact, both microchemical and instrumental methods have their own advantages and the best analysts will be familiar with all of them. Microcrystal tests for drugs are usually performed using a different technique from those conducted on inorganic ions. A small sample (many tests are sensitive enough to react with a single crystal) is placed on a slide, a drop of the reagent is placed over it and a coverslip is applied. A positive reaction is indicated by the formation of characteristic crystals that form over minutes or hours, as organic reactions generally occur over a greater period of time than inorganic ones. The determination of certain optical properties of the crystalline reaction products (birefringence, sign of elongation, extinction angle) helps to make the identifications more certain still. When used by a well-trained and experienced chemist, using positive and negative controls, the results are as accurate and reliable as those obtained from instruments. The major disadvantage is that the analyst must develop the skill to conduct and interpret these tests because the techniques are no longer taught in analytical chemistry programs at universities.

Crystal tests are useful for the identification of many other organic substances besides drugs. Specific tests for blood are based on the formation of crystals. The best known tests are those of Teichman and Takayama. In the first test a small crust of blood is heated with a mixture of potassium chloride, bromide and iodide with glacial acetic acid. The formation of brown rhombohedral crystals indicates the presence of blood. The Takayama test is based on the formation of pink crystals of pyridine hemochromogen when a little of the dried blood is warmed under a coverslip with a mixture of sodium hydroxide, pyridine and glucose.

Microchemical Analysis

Microchemical tests and techniques are best used in conjunction with sensitive analytical instruments when the problem is the identification of small amounts or single particles of complete unknowns. Chemical testing can provide useful preliminary information and can help in the interpretation of instrumental data. For example, the infrared spectrum of an impure unknown substance may be difficult to interpret in terms of the functional groups which are actually present. Well-selected microchemical tests for specific functional groups can help to answer the question. The K lines of the EDS spectra of lead and sulfur overlap and cannot be resolved by most spectrometers. Microchemical tests can rapidly distinguish between these elements. Because microchemical tests are based on chemical properties, they can be used to determine the valence state of the ions. Thus iron (II) can be distinguished from iron (III) and copper (I) from copper (II), which can be important in identifying a particular substance. Chemical reactions can also be used to distinguish between different forms of an element, for example sulfur, sulfate, sulfide, sulfite and thiosulfate, while the EDS spectrum shows only the presence of sulfur. Nitrogen species are difficult to detect and identify with certainty by most instrumental methods, but specific microchemical tests of high sensitivity are available for the ammonium ion, nitrates and nitrites.

Microchemical techniques are also important as tools for the preparation of small samples or single particles for instrumental analysis. Thus polymers often contain plasticizers and inorganic fillers. The infrared spectrum obtained from a small particle of such a polymer will be confusing, at the least. The microchemical analyst would, from experience, recognize this possibility and could devise a plan to perform the necessary separations on a microchemi-

cal scale. Thus, the particle might be placed on a polished salt plate and extracted with droplets of ether delivered from a capillary pipette with the diameter of a hair. If the droplets are drawn off to the side with a tungsten needle, before they evaporate, a crescent-shaped droplet of the plasticizer will form there and the infrared spectrum can be quickly collected. The extracted particle might now be pressed to a thin film and its infrared spectrum obtained. Inorganic spectral peaks in the particle will be easier to interpret if it is transferred to a polarizing microscope. Many inorganic fillers and pigments, such as talc, calcite, quartz and titanium dioxide, can be readily identified on the basis of their optical properties. If necessary, to complete the interpretation, the particle can be washed free of index of refraction oil and mounted on a polished beryllium plate for EDS spectroscopy to confirm the elemental composition of the fillers. This hypothetical problem illustrates the advantages of an analytical approach, based on the utilization of both microchemistry and modern instrumentation, for the analysis or identification of microscopic items of evidence.

See also: **Analytical Techniques**: Spectroscopy: Basic Principles. **Serology**: Blood Identification; Bloodstain Pattern Analysis. **Analytical Techniques**: Spectroscopic Techniques.

Further Reading

Alimarin I and Petrikova N (1964) *Inorganic Ultramicroanalysis*. New York: Pergamon.

Amelink F (1962) *Rapid Microchemical Identification Methods in Pharmacy and Toxicology*. Amsterdam: Netherlands University Press.

Benedetti-Pichler AA (1964) *Identification of Materials via Physical Properties, Chemical Tests and Microscopy*. New York: Springer.

Chamot EM (1940) *Handbook of Chemical Microscopy: Vol. II: Chemical Methods and Inorganic Qualitative Analysis*. 2nd edn. New York: Wiley.

Cheronis ND and Entrikin JB *Semicro Qualitative Organic Analysis. The Systematic Identification of Organic Compounds*, 2nd edn. New York: Interscience.

Clarke EGC (1969, 1975) *Isolation and Identification of Drugs in Pharmaceuticals, Body Fluids and Post-Mortem Material*, vol. 1 (1969) and vol. 2 (1975). London: Pharmaceutical Press.

Feigl F (1966) *Spot Tests in Organic Analysis*, 7th edn. Amsterdam: Elsevier.

Feigl F (1972) *Spot Tests in Inorganic Analysis*, 6th edn. Amsterdam: Elsevier.

Fulton CC (1969) *Modern Microcrystal Tests for Drugs. The Identification of Organic Compounds by Microcrystalloscopic Chemistry*. New York: Wiley-Interscience.

Geilman W von (1954) *Bilder zur Qualitativen Mikroanalyse Anorganischer Stoffe*. Weinheim: Verlag Chemie.

Jungreis E (1997) *Spot Test Analysis Clinical, Environmental, Forensic and Geochemical Applications*, 2nd edn. New York: Wiley-Interscience.

Palenik SJ (1979) Microchemical reactions in particle identification. In: McCrone WC, Delly JG and Palenik SJ (eds) *The Particle Atlas*, 2nd edn, pp. 1174–1175. Ann Arbor: Ann Arbor Science.

Schneider FL (1964) *Qualitative Organic Microanalysis. Cognition and Recognition of Organic Compounds*. New York: Academic Press.

Microscopy *see* **Analytical Techniques**: Microscopy. **Microchemistry**.

Microtrace *see* **Dust**.

MODUS OPERANDI

B E Turvey, Knowledge Solutions LLC, Watsonville, CA, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0575

Introduction

Modus operandi (MO) is a Latin term that means ‘a method of operating’. It is used by law enforcement agencies to refer to a criminal’s pattern of behavior,

or his or her way of committing crime. It is by understanding a criminal's MO behavior that most of those in law enforcement investigate and link criminal cases. This investigative approach has evolved from the use of criminal informants and highly experienced detectives, into the additional use of criminal database systems. The use of MO and criminal databases to link cases is but one of several important investigative methods available to law enforcement, and is separate from the concept of offender '*signature*'.

Understanding Criminal Behavior

Law enforcement has long held to the belief that understanding the methods criminals use to commit crime is the best way of searching for, and ultimately apprehending, them. This has traditionally required that detectives be a living encyclopedia of criminal cases and criminal behaviors, or that they learn to utilize the knowledge and experience of known criminals to inform their investigations. A strong belief in this way of approaching the process of criminal investigation was well demonstrated in France, in 1817, when a former convict named Eugene Vidocq, who had been working as a police spy, was assigned by the government to form a *Brigade de Sûreté*. He organized and led this group of detectives, mostly former criminals themselves, as it grew from a mere four to 28 in number. Vidocq and his detectives were paid according to the number of criminals that they apprehended. Within their first year, the group had made more than 750 arrests. This led some to believe that Vidocq and his detectives were a perfect solution to the local criminal problem; they understood how criminals operated, had insights into their habits and methods of operation, and were putting that knowledge to work for the good of the state. However, others suspected that Vidocq and his detectives committed many of the crimes themselves and then framed known criminals or undesirables in order to close out the cases. This suspicion was never proven, however, and Vidocq enjoys a largely favorable historical place as the first, and very successful, chief of the French *Sûreté Nationale*.

Whether or not Vidocq was a master detective or merely continued his criminal career through the Sûreté, a philosophy of criminal investigation emerged. To understand criminals, to develop competent investigative strategies, to link their crimes and to successfully apprehend them, detectives needed to understand the particular methods criminals used to commit their crimes. This is an investigative philosophy that survives on an international level, in one form or another, to the present day.

Definitions

Modus operandi has been defined as the actions taken by a criminal to perpetrate a crime successfully. A criminal's MO is comprised of learned behaviors that can evolve and develop, as they become more sophisticated and more confident. It is through analyzing MO behaviors that law enforcement agencies most commonly investigate and link crimes to a specific offender. A criminal's MO behavior is functional in nature. It generally serves any of three purposes for the offender:

- to protect identity
- to ensure success
- to facilitate escape

General types of MO behaviors include, but are not limited to:

- offense location selection (i.e. in a public park, on a school campus, or in a victim's residence, etc.);
- involvement of a victim during a crime;
- use of a weapon during a crime;
- use of restraints to control the victim during a crime;
- offender precautionary acts (i.e. wearing a mask, gloves, covering the victim's eyes during an attack, wearing a condom during a rape, forcing the victim to bathe after a sexual attack, etc.);
- offender transportation to and from the crime scene (i.e. use of a bicycle, use of a motorized vehicle, walking, etc.).

The criminal's MO, evidenced by MO behavior, is not the same thing as a criminal's *motive*, which is their reason for committing the crime. A criminal's motives are evidenced by *signature behaviors* that suggest overall *signature aspects*, or motivational aspects to the crime. Signature behaviors are those that satisfy the offender's emotional and psychological needs. They are often specialized behaviors, typically not necessary for the completion of the crime, and tend to show less evolution across offenses than MO behavior.

Influences on Modus Operandi

A criminal's MO behavior is learned, and therefore dynamic and malleable. This is because MO behavior is affected by time, and can change as the criminal discovers that some of the things done during a crime are more effective than others. Criminals can subsequently recognize these effective actions, repeat them in future offenses, and become more skillful, refining their overall MO. However, behavior may also change due to a criminal's deteriorating mental state, due to

the influence of controlled substances, and/or due to increased confidence that law enforcement will not successfully apprehend them. These things may cause a criminal's MO to become less skillful, less competent and more careless.

Common ways that criminals can learn how to commit crime more skillfully are by gaining more experience, building confidence through success and/or having more contact with the criminal justice system. Being arrested just once may teach an offender an invaluable lesson about how to avoid detection by law enforcement in the future. Further still, and with some great irony, a prison term in the United States is referred to by some, in both law enforcement and the criminal population, as 'going to college'. This is because younger and less experienced offenders have the opportunity in prison to network with older and more experienced offenders who have already accumulated a great deal of criminal knowledge. Subsequently, a prison term of only a few years has the potential to advance an offender's skill level far beyond their original MO. Once released, such offenders may take their 'education' and embark on criminal enterprises that before would have been beyond their ability.

Criminals can also seek out knowledge directly, just like anyone else, without having to spend time with other, more experienced, criminals. They can learn from the things that they see or read, which include items ranging from those in the media to educational and technical materials. For example, a rapist may commit five different attacks in a single region. The attacks may go unconnected until DNA results come back and demonstrate that the rapes were more than likely committed by the same offender. If the media publishes a headline that reads 'Serial rapist linked to five attacks by DNA!', the rapist may alter his MO behavior to prevent law enforcement from linking future cases. He may do so by making temporary changes, such as using a condom during any future rapes, or he may decide to make a more permanent change and undergo a vasectomy. Either way, the rapist may make a conscious attempt to prevent the transfer of a particular type of evidence, based on what he has learned from the media coverage of the case or other similar cases.

However, MO behavior does not always evolve to become more competent as offenders progress through their criminal careers. Owing to a deteriorating mental state, the use of controlled substances, or increased confidence that law enforcement will not successfully apprehend them, offenders' MO behaviors can evolve over time to a less competent and less skillful level than when they first began. For example, the American serial murderer Theodore (Ted) Bundy, who killed at least 30 victims across five states between 1973 and 1978, began his criminal career with a very competent, very well thought out MO. He was polite and friendly, extremely mobile and often approached his victims in some manner as to appear helpless or weak, and essentially nonthreatening. He sometimes accomplished this by presenting himself as a motorist who needed assistance with a disabled vehicle, and would often wear his arm in a sling. He also tended to select teenage females as victims, stalking them and selecting a disposal site for their bodies well in advance of committing an actual crime. But his MO deteriorated remarkably over time. After being incarcerated and then escaping on two separate occasions in Colorado, he made his way to Florida. He began to drink heavily, and he began to involve the bodies of his victims in rituals (*signature behavior*), including keeping the body for days at a time after death. There was also evidence that Bundy shampooed some of his victim's hair, and applied make-up to their corpses, rather than disposing of them immediately. In short, he began to leave more and more evidence behind, engaged in fewer precautionary acts, and became involved in more ritual behavior. His victim selection also changed; he chose his last victim, a 12-year-old female student from Florida, totally by virtue of her availability. This was a marked departure from his previous MO behavior of carefully stalking victims in advance and selecting victims who were in their late teens and early twenties.

Understanding how MO can change over time, and the types of things that can influence those changes, has become an important part of using MO as an effective investigative tool.

Linkage Blindness

The MO behavior demonstrated by a criminal can play an important role in law enforcement's linkage of related crimes. However, because MO behaviors can change over time, the tendency of law enforcement to rely solely upon an offender's MO for investigative strategy can lead to what has been termed '*linkage blindness*'; the failure to recognize a pattern that links one crime with another crime in a series of cases.

Specifically, there are three very important factors that can act individually or in concert to cause linkage blindness:

- the tendency for law enforcement to rely solely on MO behaviors like victim type, weapon selection and location type as a basis for case linkage;
- the possibility that one predatory offender is operating in or near the same general area as another, confusing law enforcement efforts;

- interpersonal or interagency conflicts, which can lead to communication breakdowns and a lack of information sharing.

Crime Linkage Systems: VICAP and VICLAS

If you take a picture of a person who has committed a crime, then that person can be identified if he or she commits another crime later on. This fact was not lost on law enforcement as they began to incorporate photography into the criminal investigative process during the last half of the 1800s. As the sheer number of photographs of criminals began to grow, the need for a system of classification became apparent. The first classification systems adopted by law enforcement for their growing photographic databases were oriented towards MO behaviors. Photographs were sorted by virtue of the basic type of crime, and perhaps methods, used to commit the crime. Pickpockets who operated at racecourses went into one pile, while burglars who first 'cased' the homes they intended to rob by posing as salesmen went into another.

By the early 1900s, the contents of photographic databases had grown to substantial numbers and methods of classification had become more sophisticated and more competently crossreferenced. Criminal photographs were no longer simply classified by basic MO behaviors. Classification had evolved into advanced databases of criminal activity that were subclassified by factors such as method of entry into the crime scene, materials brought to commit the crime (i.e. flashlight, rope, lock picks, etc.), type of weapon used and physical description.

Today, those at the forefront of developing law enforcement investigative techniques understand that there are a number of important, but often ignored, ways to link cases, apart from direct physical evidence and witness or victim statements. These include:

- *Modus operandi*: similarities between actions taken by an offender that are necessary for the successful completion of the crime.
- *Signature: similarities between actions taken by an offender that are unnecessary for the successful completion of the crime; actions that suggest a psychological or emotional need.*
- *Victimology*: similarities or connections between victims.
- *Wound pattern analysis*: a review of the autopsy or hospital protocols for a victim in an effort to find similarities between the nature and extent of injuries sustained during an attack.
- *Geographic region or location*: offenses that have occurred in the same area, or the same type of area.

There is a large amount of published information regarding the limitations of using MO behavior as a sole basis for linking solved and unsolved criminal cases, and the importance of understanding the concept of offender signature. However, even the most modern law enforcement case linkage initiatives tend to focus on understanding MO behavior alone.

Violent Criminal Apprehension Program (VICAP)

The core of these ideas was brought into modern offender databasing by the late Pierce Brooks, an American law enforcement officer. Brooks joined the Los Angeles Police Department in 1948 and served in the vice, narcotics and homicide divisions. In 1958, Brooks was assigned two homicides that, although clearly unrelated, appeared to be the work of individuals who had killed before. He looked for resources to assist him in linking either of his homicides to any other cases in other parts of the country. There was nothing like that available on a national level. So Brooks used the only national information system available to him – he went to the local library for hours at a time and sifted through newspapers from around the country in the hopes of finding stories that related cases with similar MO behavior.

Brooks quickly came to realize that what was needed was a national database and information center that collected information on the MO of those killers that crossed jurisdictional boundaries. As early as 1958, he also knew that a computerized database of such information was the most effective way to make it nationally accessible. His technique of searching through library newspaper archives for related cases, born out of extreme necessity, was the basis for what was to become the Federal Bureau of Investigation's *Violent Criminal Apprehension Program* (VICAP).

In 1983, Brooks testified before Congress about the possibility that unsolved murders around the country might be attributed to anonymous serial murderers. He explained that the only way to investigate the crimes and link the cases to apprehend these individuals was to put all of the information about each case into a computer database system that everyone in law enforcement could have access to. In 1984, with Brooks' help, the FBI's National Center for the Analysis of Violent Crime (NCAVC) was officially established. VICAP went online in 1985, with Brooks as the first program manager. Supervisory Special Agent Robert K. Ressler followed Brooks as the first agent program manager after Brooks terminated his FBI consultation.

VICAP remains a nationwide data information center, specifically designed for collecting, sorting and analyzing solved and unsolved homicides, as

well as missing persons cases where there is a strong possibility of foul play. According to the FBI, VICAP's mission remains to facilitate cooperation, communication and coordination between law enforcement agencies and provide support in their efforts to investigate, identify, track, apprehend and prosecute violent serial offenders.

Law enforcement agencies can submit their unsolved and solved cases for comparison to those in the VICAP system via a handwritten 15-page report form. The VICAP Crime Analysis Report details administrative, victim and offender information, including MO, as well as information about location, forensic examinations and potentially related cases. The purpose of submitting the information is so that VICAP analysts can detect signature aspects and similar patterns of MO, and determine whether or not any cases in the current database may be linked.

VICAP suffers greatly from underutilization by law enforcement. The primary complaint tends to be that the form is too long and too inconvenient to fill out in every case. End users of VICAP further complain that the database is too small, and that it should contain information regarding other types of interpersonal crime as well, such as rapes and sexual assaults. To augment the limited number of cases in the database, VICAP analysts have taken a page from the book of Pierce Brooks, and have populated it with information on solved and unsolved cases obtained from newspaper articles. Another weakness of VICAP, and similar statewide databases, is that analysts still predominately use MO behaviors to make initial case linkages. This essentially ignores the many other ways of avoiding linkage blindness that have proven more robust, such as offender signature.

Although VICAP has not been significantly changed or modified since its inception in 1985, VICAP 2000 has been implemented, which expands the data base somewhat.

Violent Crime Linkage Analysis System (VICLAS)

Following research into the FBI's VICAP system, the Canadian law enforcement community developed what was referred to as the Major Case File (MCF). It was Canada's first attempt to link homicide cases on a national level. Much like the VICAP report, investigators filled out their forms and submitted them to regional MCF analysts, who would manually input the data into the system. The MCF database was searchable by key words and phrases, and detectives would attempt to link offenses largely by MO behaviors. Despite the approximately 800 cases in the MCF database, no linkages had been made with the system as late as 1990.

In 1991, a Royal Canadian Mounted Police inspector, Ron McCay, who had recently received training through the FBI's criminal profiling fellowship program in Quantico, Virginia, helped to develop a new linkage system for Canada. It was to become the Violent Crime Linkage Analysis System (VICLAS). Its development was the result of analyzing not only the FBI's VICAP database but other statewide linkage systems in the United States as well.

A far more ambitious and robust tool than the FBI's VICAP, VICLAS attempts to input information from the following sources:

- all solved or unsolved homicides and attempted homicides;
- solved and unsolved sexual assaults;
- missing persons, where there is a strong possibility of foul play and the victim is still missing;
- unidentified bodies, where the manner of death is known or suspected to be homicide;
- all nonparental abductions and attempted abductions.

An emphasis was also placed on the training and qualification of VICLAS analysts as an integral part of the system. Submissions still occur via a questionnaire that is filled out by an investigator and is sent to one of ten VICLAS centers servicing the various Canadian regions. However, there are numerous quality assurance reviews in place, and the VICLAS analysts may return forms to detectives if something is questioned. VICLAS analysts are trained to conduct structured queries of the system based on their own expertise, and arrive at potential linkages based on their own analysis of victimology, offender behavior, MO and forensic information. They are also trained to understand the concept of offender signature. Regardless of the results of their analysis, investigators in the field are charged with ultimately confirming or rejecting the links, based on the substance of their own investigation.

The Canadian VICLAS initiative has become a popular model for law enforcement agencies involved in violent crime case linkage. Similar systems to VICLAS, though not directly connected, have been adopted in other countries, including Australia, Austria, Belgium, Holland and the United Kingdom. This speaks to the trend in the international law enforcement community to admit both that they have serial crime, and that an understanding of the relationship between MO and other elements that comprise criminal patterns is necessary to successfully link and investigate them.

See also: **Crime-scene Investigation and Examination**: Scene Analysis and Reconstruction. **Offender Signature**.

Further Reading

Barbaree H, Laws D and Marshall W (1990) *Handbook of Sexual Assault: Issues, Theories, and Treatment of the Offender*. New York: Plenum Press.

Bross H (1924) *Criminal Investigation.* (English Translation). London: Sweet and Maxwell.

Burgess A and Hazelwood R (eds) (1995) *Practical Aspects of Rape Investigation: A Multidisciplinary Approach*. New York: CRC Press.

Burgess A, Groth A and Holmstrom L (1978) *Sexual Assault of Children and Adolescents*. New York: Lexington.

Burgess A, Douglas J and Ressler R (1988) *Sexual Homicide: Patterns and Motives*. New York: Lexington.

Burgess A, Burgess A, Douglas J and Ressler R (1997) *Crime Classification Manual*. San Francisco: Jossey-Bass.

Canter D and Alison L (1997) *Criminal Detection and the Psychology of Crime*. London: Ashgate.

Egger S (1997) *The Killers Among Us*. Englewood Cliffs, NJ: Prentice Hall.

Geberth V (1996) *Practical Homicide Investigation*, 3rd edn. New York: CRC Press.

Groth AN (1979) *Men Who Rape: The Psychology of the Offender*. New York: Plenum Press.

Holmes R and Holmes S (1996) *Profiling Violent Crimes: An Investigative Tool*, 2nd edn. Thousand Oaks, CA: Sage.

O'Hara H (1970) *Fundamentals of Criminal Investigation*. Springfield: Charles C. Thomas.

Turvey B (1999) *Criminal Profiling: An Introduction to Behavioral Evidence Analysis*. London: Academic Press.

Motor Vehicles *see* **Accident Investigation**: Airbag Related Injuries and Deaths; Driver Versus Passenger in Motor Vehicle Collisions; Motor Vehicle.

Murder *see* **Crime-scene Investigation and Examination**: Collection and Chain of Evidence; Contamination; Criminal Analysis; Fingerprints; Packaging; Preservation; Recording; Recovery of Human Remains; Scene Analysis and Reconstruction; Suspicious Deaths.

N

NURSING

C M Dougherty, MTF, Edison Corp. Cte, Edison, NJ, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0744

Definitions and Role Delineation

The term 'forensic' is derived from the Latin *forensis*, which defined a public debate in the market plaza, i.e. forum. Eventually the term came to be used to describe debates in courts of law. Forensic sciences is a collective term for various disciplines that have components involving interfaces with law, e.g. forensic engineering, forensic medicine, forensic pathology, forensic psychiatry and, a recent addition, forensic nursing. As defined by the Scope and Standards of Forensic Nursing Practice, forensic nursing is:

> The application of forensic science combined with the bio-psychological education of the registered nurse, in the scientific investigation, evidence collection and preservation, analysis, prevention and treatment of trauma and/or death related medical issues. The forensic nurse functions as a staff nurse, nurse scientist, nurse investigator or as an independent consulting nurse specialist to public and/or private operatives and/or individuals in the medical-legal investigation of injury and/or death of victims of violence, criminal activity and traumatic accidents. The forensic nurse provides direct and indirect services to individual clients, consultation services to nursing, medical and law related agencies, as well as providing expert court testimony in areas encompassing evidence collection, preservation and analysis, questioned death investigation processes, adequacy of services delivered and specialized diagnosis of specific conditions related to nursing practice.

The pictorial model of the theoretical foundation of forensic nursing portrays a nursing professional whose nursing science background must incorporate forensic science and criminal justice principles (**Fig. 1**). It requires a forensic nurse to deliver care within health care institutions to the alleged victim, perpetrator and significant others, diagnosing specific conditions related to nursing. He or she is further expected to understand human behavior as it pertains to the acts of, and response to, crime and violence in the light of social sanction.

At the top of the triangle are three areas from which the knowledge base of forensic nursing is derived: nursing science, forensic science and criminal justice. The model assumes that the nurse embraces the philosophy and principles of these three disciplines and shares a mutual responsibility with law enforcement agencies and courts in protecting the legal, civil and human rights of crime victims. One additional tenet of this role is the recognition and protection of the constitutional rights of the suspect or perpetrator of criminal acts. The lower left corner represents the social sanctions, human behavior and crime and violence, recognizing that each society has its unique aggregate of crime and violence, dictated by its dynamic social constraints. The lower right corner portrays the emerging discipline of the forensic nurse and the locus of its practice domain. At the center are the scales of justice overlaid with a caduceus. The bundle of public service forms the base. The flame of nursing is noted at the apex of the triangle, denoting enlightenment in a new field of practice. The symbol is enclosed by an interlocking circle that emphasizes the multidisciplinary coordination and cooperation among sociology, criminology, education, culture, politics and healthcare systems.

Historical Developments

Clinical forensic medicine is an old discipline and its roots in the United Kingdom and other parts of the world can be traced back for over 200 years. Clinical forensics, a relatively new concept focusing on living victims, as opposed to the dead, may be attributed to Harry C. McNamara, Chief Medical Examiner for Ulster County, New York. In 1987, McNamara defined clinical forensic medicine as 'the application of clinical medicine to victims of trauma involving the proper processing of forensic evidence'. This definition stresses the importance of healthcare providers

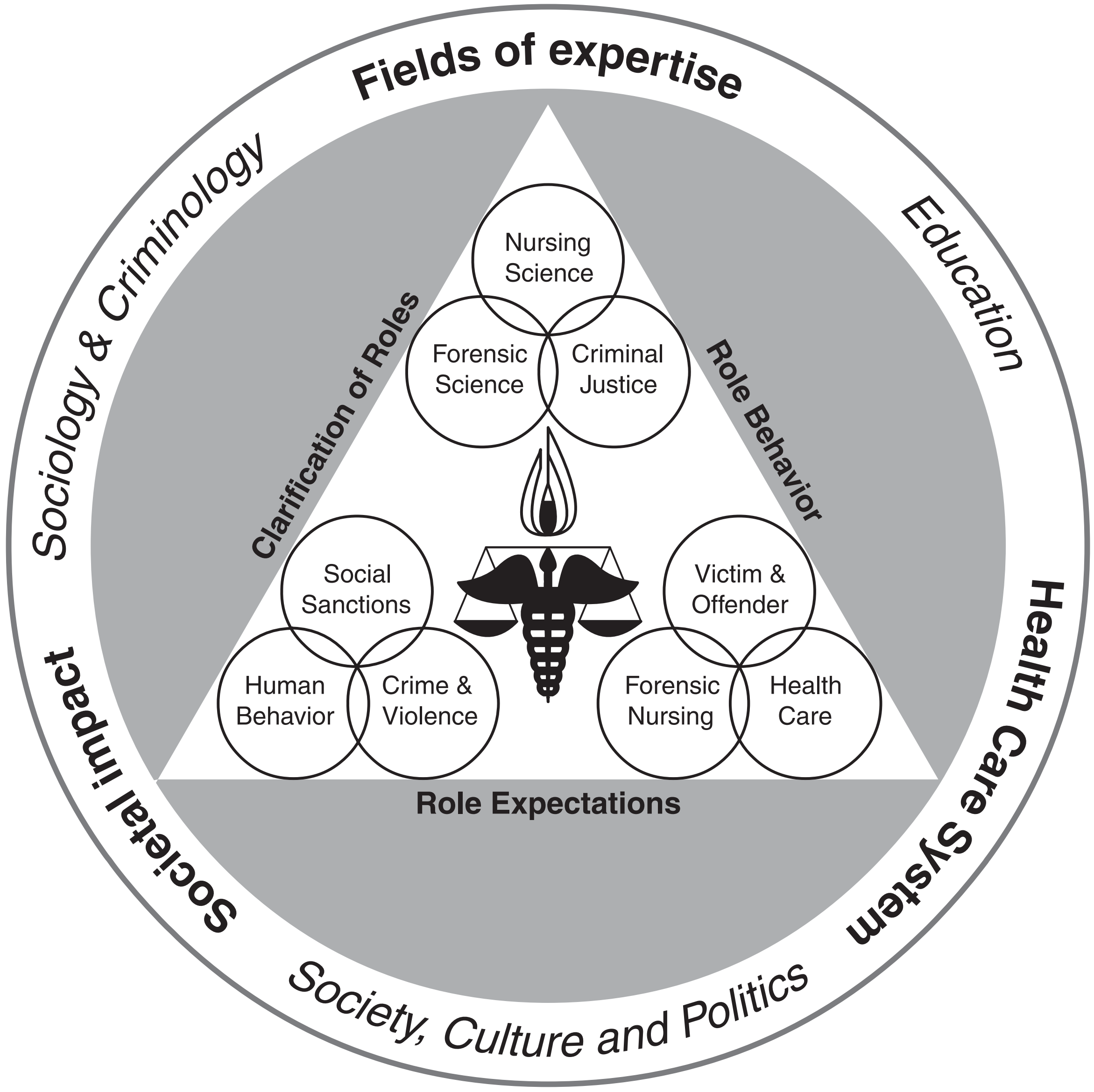


Figure 1 Conceptual model of forensic nursing practise.

being aware of evidentiary materials and legal issues associated with their patients/clients. Although the concept is new to the United States, it has been practiced for many years in other countries.

The roots for the development of forensic nursing in the United States can be traced to the initiatives of nurses who were engaged in the examination and care of sexual assault victims. In Minneapolis, Minnesota during the summer of 1992, spearheaded primarily by these sexual assault nurse examiners, the vision of an organization encompassing and defining the broad spectrum of forensic nursing, both in the United States and beyond, crystallized in the founding of the International Association of Forensic Nursing (IAFN). In 1995 the American Nurses Association (ANA) Congress of Nursing Practice met and approved a request made by the IAFN to recognize forensic nursing as a designated nursing specialty.

The ANA, the largest professional nursing organization in the United States, provided credibility and support to nurses practicing nontraditional roles in traditional settings. The official recognition of forensic nursing paved the way for greater acknowledgment and respect among the other well-established

disciplines of forensic sciences worldwide. In the last two decades, forensic nurses have been accepted for full membership within the Academy of Forensic Sciences, and even to fellowship status.

Organizational Associations: International Association of Forensic Nurses

There are various groups of organized nurses that have embraced forensic-related nursing issues; for example, the American Association of Legal Nurse Consultants, American Academy of Nurse Practitioners, Emergency Nurses Association, Sigma Theta Tau International, American Correctional Health Services Association and the ANA. However, until the formation of the IAFN (home office in Thorofare, New Jersey), there was no organization expressly serving the basic needs of forensic nurse specialists. IAFN acts as an educational and informational resource for forensic nurses throughout the world. Its nearly 2000 member family provides an opportunity for forensic nurses to conduct the business of subspecialty development and to network regarding role evolution, educational programming, practice areas and research endeavors. Approval for local branch development was granted in 1998, with the intent for these branches to:

- provide a vehicle for professional and public education;
- foster interaction and networking between professionals and groups of people who share common interests or concerns;
- encourage resource and information sharing;
- have an increased ability to affect local policies and regulations;
- provide a base for leadership development.

The IAFN coordinates official publications, distributes a quarterly newsletter, *On the Edge*, and conducts an annual scientific assembly. The early assembly programs were dominated by plenary speakers representing forensic pathology, criminology, criminalistics and other associated disciplines. However, in recent years there has been an obvious trend toward increasing emphasis on the nursing component of the discipline. The Sixth Scientific Assembly, held in September 1998 in Pittsburgh, Pennsylvania, featured a roster of primarily nurse experts. The scientific paper presentations emphasized nursing research findings and nursing case studies. The assembly was attended by hundreds of nurses from the United States and several other countries. It is the largest scientific program expressly designed for the educational needs of forensic nurses in the world. The first international meeting will be in Calgary, Alberta, Canada in 2000.

Additionally, the IAFN, through its Ethics Committee, offers a venue for forensic nurses to resolve ethical issues involving professional practice matters. Created in 1994, the IAFN's Committee on Ethics developed the Forensic Nurse's Code of Ethics and presented it to the membership the following year. The code, modeled after that of the American Chemical Society, was tailored to address the unique aspects of forensic nursing.

IAFN Forensic Nurse's Code of Ethics

The International Association of Forensic Nurses expects its members to adhere to the highest standards of ethics. Forensic nurses have professional obligations to colleagues, to science, and to the public, especially to those members of the public who are demonstrably disadvantaged. Accordingly, the International Association of Forensic Nurses expects its members and associate members to abide by its Code of Ethics as a condition of initial and continued membership.

Responsibility to the public and the environment

Forensic nurses have a professional responsibility to serve the public welfare, especially most disadvantaged citizens, and to further the cause of science and justice. Forensic nurses should be actively concerned with the health and welfare of the community at large. Forensic nurses must understand and anticipate the environmental consequences of their work and the work of others in their communities. Forensic nurses must be prepared to stand up and oppose environmental pollution and other environmental degradation. Public comments on scientific matters should be made with care and precision, devoid of unsubstantiated claims, exaggeration, and/or premature statements.

Obligation to science

Forensic nurses should seek to advance nursing and forensic science, understand the limits of their knowledge, and respect the truth. Forensic nurses should insure that their scientific contributions, and those of their collaborators are thorough, accurate and unbiased in design, operationalization, and presentation.

Care of the profession

Forensic nurses should remain current with developments in their field, share ideas and information, keep accurate and complete records, maintain integrity in

all conduct and publications, and give due credit to the publications of others. Conflicts of interest and scientific misconduct, such as fabrication, falsification, slander, libel and plagiarism, are incompatible with, and a violation of this Code.

Forensic nurses, as employers, should promote and protect the legitimate interests of their employees, perform work honestly and competently, fulfill obligations, and safeguard proprietary information. As employees and managers, forensic nurses should treat subordinates with respect for their professionalism and concern for their well being, and provide them with a safe, congenial working environment, fair compensation, and proper acknowledgments of their scientific contributions. Forensic nurses should regard the tutelage of students as a trust conferred by society for the promotion of the student's learning and professional development. Each student should be treated respectfully and without exploitation. Forensic nurses should treat associates with respect, regardless of the level of their formal education, and encourage them, learn with and from them, share ideas honestly, and give credit for their contributions.

Fidelity to clients

Forensic nurses should serve clients faithfully and incorruptibly, respecting confidentiality, advising honestly, and charging fairly.

Training and Education

Forensic nursing practice encompasses a vast body of knowledge and skills derived from theoretical bases in nursing, medicine, criminal justice, law enforcement, psychology, social sciences, chemistry, environmental sciences and law. Due to the dearth of formal curricula for forensic nursing preparation, nurses have been forced to search for education and training experiences in nontraditional programming. Some nurses enrolled in criminal justice programs, death investigator courses and internships in the medical examiner/coroner offices, where their nursing education was often overlooked and undervalued. Although investing time and money in forensic science courses based outside the realm of nursing is generally still considered by many forensic nurses to be a valuable learning experience, there is an increasing and vocal need to see formal nursing education respond to the nursing of the twenty-first century. To date there are at least three graduate programs in the United States that offer a Master's Degree in Forensic Nursing: Fitchburg State College in Fitchburg, Massachusetts; the University of Colorado – Beth El, Colorado Springs, Colorado; and Gonzaga University, Spokane, Washington. There are continuing education and/or selected undergraduate or graduate courses at the University of Texas, the University of Virginia, the University of Massachusetts, and a unique undergraduate certificate program in Lincoln, Nebraska. This program, affiliated with a local university, is a 20-credit course preparing nurses, social workers, law enforcement and mental health workers to integrate necessary forensic science skills into their practices. The first forensic Internet distance delivery course was written and instructed by forensic clinical nurse educator at Mount Royal College, Calgary, Alberta, Canada.

In an evolving professional discipline, there is a fundamental requirement to identify the core knowledge and skills essential for practice, and subsequently certification or licensure. In 1998, the IAFN appointed a committee to define essential knowledge and skills necessary for practice within the discipline. The original IAFN core curriculum was drafted at a conference held in Indianapolis, Indiana in August 1998. The peer review, publication and utilization of this document is a current initiative of the organization, working in concert with the ANA and other pertinent affiliates. In addition to basic certification for forensic nursing, there is ongoing study regarding subspecialty certification within the broader forensic nursing field. The Sexual Assault Nurse Examiner and Death Investigation Councils of the IAFN have begun preliminary consideration for additional credentialing processes.

Forensic Nursing and the Healthcare Community

The incidence of injury and death as a result of violence is a well-known public health issue. The field of forensic nursing represents an alliance of nurses as healthcare professionals, law enforcement and the forensic sciences in an effort to better identify and resolve issues of violence. The evolution of a new area of clinical practice for nurses will assist in and provide another approach to the management of criminal and interpersonal violence.

Healthcare personnel have a specific mandate to report cases of human abuse and neglect to proper authorities. The Joint Commission on the Accreditation of Hospitals in the United States has developed Standard PE.1.9, which addresses victims of alleged or suspected abuse or neglect. The Commission suggests that appropriate nursing care cannot be rendered 'unless suspected or alleged victims are identified and assessed. Assessing the care needs of the patient is aided by using established criteria that

focus on objective evidence that is observable (as opposed to action on allegation alone) to identify possible victims of abuse throughout the organization.' Nurses have been mandated to take responsibility for the recognition, collection and preservation of forensic evidence.

Forensic Nursing Roles: Practice Arenas

A former Surgeon General of the United States described the social and legal response to victims of trauma as 'late and inadequate'. Nurses are often the first caregivers to see victims of trauma, which most logically makes them key players in the recognition, collection and preservation of forensic evidence. Nurses function in a variety of roles to provide forensic services to clients. These roles may be divided into two categories: clinical forensic nursing, which involves caring for the living, and postmortem forensics. Clinical forensic nurses may care for victims of violent crime and human assault, domestic violence, substance abuse, automobile/pedestrian accidents, suicide attempts, occupational-related injuries, disputed paternity cases, medical malpractice, psychological abuse, drug and food tampering, environmental hazards, nursing home neglect/abuse, and communicable diseases posing threats to public safety.

The role of the clinical forensic nurse may be established within a healthcare or social facility, or in the community when there are gaps between healthcare and the criminal justice systems in regard to services for victims of trauma and human abuse or violence. The nurse who practices in this arena works with law enforcement officers, paramedical personnel, physicians, social workers, attorneys and others who provide services to victims. Although nurses have often functioned as liaisons between survivors and nonsurvivors of traumatic incidents and the legal system, it is only recently that official recognition has come to the role in the United States, and subsequent greater emphasis and attention to the role internationally.

Those nurses working as death investigators or coroners practice what is termed postmortem forensics or nonliving victim forensics.

There are several subspecialties of forensic nursing (**Table 1**), two of which are described in further detail below.

Table 1 Subspecialties in forensic nursing

Title	*Practice area*	*Forensic function*
Sexual assault nurse examiner	Hospital, woman's center, free standing clinic, police department, prosecutor's office, independent contractor	Conducts a forensic examination of victims of rape, to include physical examination with written documentation and often photographic as well as culposcopic record
Emergency department/accident ward nurse	Emergency department/accident ward	Identifies victims of violence or abuse and conduct a forensic examination. Provides appropriate referral to available services
Anatomical gift nurse	Hospital	Assists families in exercising their option and legal right to organ and tissue donation in conjunction with the medical examiner's office
Epidemiology nurse	Hospital, nursing home, school, sexually-transmitted disease clinic, any public gathering place	Investigates public health issues to help prevent the spread of communicable diseases
Geriatric nurse	Nursing homes, public health nursing offices	Acts as an advocate for the elderly; investigates conditions and treatment of the elderly
Legal nurse consultant	Attorney's office, independent contractor	Acts as a consultant on medicolegal issues, particularly as they relate to nursing care
Corrections nurse	County/state prisons, jails	Cares for the convicted criminal's physical health needs while he or she is incarcerated
Psychiatric forensic nurse	County/state prisons, jails, hospitals, clinic setting	Assesses, evaluates and treats the defendant prior to criminal hearing or trial
Medicolegal death investigator	Medical examiner/coroner's office	Investigates deaths at the direction of the medical examiner/coroner

Sexual assault nurse examiner

Perhaps the most easily recognizable and accepted forensic nurse is the sexual assault nurse examiner. Programs throughout the United States can claim 10 years and more of service to their communities. Many sexual assault nurse examiner programs were the result of victims' advocacy groups looking for a more timely and compassionate manner in which to deal with victims of sexual assault. Although program sites, team composition and training requirements may vary slightly, the mission of the sexual assault nurse examiner is to conduct a forensic examination of victims, and sometimes the perpetrators.

Clinical forensic nurse (emergency department)

The duties of the specialist forensic nurse attached to the emergency department or accident ward are as follows:

- Consulting with the public and/or private organizations and/or individuals in the investigation of medicolegal-related trauma and/or questioned deaths in the Emergency Department.
- Notifying family members or significant others of deaths and providing immediate crisis intervention.
- Informing law enforcement agencies and other appropriate authorities of circumstances surrounding trauma deaths that impact medicolegal actions.
- Providing information and advice for medical personnel relating to organ/tissue donation and interfacing with the victim's family members in requesting anatomical gifts.
- Serving as a liaison between the Emergency Department and Medical Examiner's Office and law enforcement.
- Representing the hospital as expert witness in civil and criminal cases.
- Maintaining current knowledge of case law, legal evidence collection procedures, and a working knowledge of law enforcement and medical investigative responsibilities.
- Participating in the educational programs of the hospital that have forensic implications.
- Consulting with the Emergency Department and other hospital departments/services as requested.
- Providing sexual assault examinations as an expert practitioner in forensic gynecology.
- Developing and maintaining standards, policies and procedures, and activities/programs essential to the forensic application to traumatic injuries in nonfatal assaults, accidents or self-inflicted wounds.
- Applying knowledge and skills of biomedical/medicolegal investigation in the immediate nursing interventions and follow-up care of victims of violence.

See also: **Clinical Forensic Medicine**: Sexual Assault and Semen Persistence. **Education**: An International Perspective. **Ethics**.

Further Reading

Barber J (1991) Frontiers and challenges in critical care: foreword. *Critical Care Nursing Quarterly* **14**(3):V.

Duval J (1995) Role of the forensic nursing specialist in an urban trauma center. In: *Proceedings of the 47th Annual Meeting of the American Academy of Forensic Sciences*, Seattle, p. 101, abstract D44.

International Association of Forensic Nurses (1998) *Guide to Organizing a Chapter of IAFN*. Thorofare, NJ: IAFN.

International Association of Forensic Nurses (1998) *Membership Directory*. Thorofare, NJ: IAFN.

Joint Commission on Accreditation of Health Care Organization (1997) *AMH Accreditation Manual for Hospitals*. Oakbrook Terrace, IL: JCAHCO.

Koop C (1989) President and Surgeon General condemn violence against women: a call for new attitudes, program. *National Organization for Victims' Assistance Newsletter* **13**: pp.

Lynch VA (1991) Forensic nursing in the emergency department: a new role for the 1990s. *Critical Care Nursing Quarterly* **14**:69–86.

Lynch VA (1995) Clinical forensic nursing: a new perspective in the management of crime victims from trauma to trial. *Critical Care Clinics of North America* **7**:489–507.

O

ODONTOLOGY

J G Clement, Foundation Chair of Forensic Odontology, University of Melbourne, Melbourne, Australia

Copyright © 2000 Academic Press

doi:10.1006/rwvi.2000.0577

Definitions

In the context of this encyclopedia 'odontology' is synonymous with 'forensic dentistry' and 'forensic odontostomatology'. None are really adequate terms for what is today one of the fastest emerging specialties within the broader subject of dental science. As all of the somewhat limited definitions above imply, 'forensic odontology' is practiced by someone who understands the significance of the conjunction of the law and dentistry and can therefore explain the complexities and subtleties of dental evidence to the courts. This capacity to report on findings and then progress to express an opinion relies upon the courts affording expert status to the witness. In the adversarial systems of justice, this may require the expertise to be established on each occasion to the satisfaction of the judge from the outset. In a practical sense, it is the combination of education, training and experience appropriate to the case currently before the court that is scrutinized and so track record has a major part to play in the decision.

The forensic odontologist not only has to be an experienced practitioner of clinical dentistry but someone who is also able to observe, record, gather, preserve and interpret dental evidence, and who is then able to make it meaningful within a legal context. In common with all good expert witnesses this requires special skills, such as the ability to communicate clearly and simply by word and in written reports.

The Nature of the Work

In Victoria, Australia, a state of approximately 4 million people, the casework for the last decade is roughly typical of that experienced in many modern, urbanized, industrialized societies. In this period approximately 20 000 coronial autopsies were conducted. Of the 504 cases (2.5% of all autopsies) needing forensic dental expertise, 463 (92%) required the establishment or corroboration of identity. In most cases the injuries to, or decay of, the body precluded visual identification. Incineration of remains was the most common cause of disfigurement (169 cases, 37%) followed by decomposition (129 cases, 28%), severe trauma (64 cases 14%) and skeletalization (40 cases, 9%).

Therefore, while forensic odontology sometimes encompasses cases dealing with ethical issues, and events relating to claims of negligence or malpractice by dentists and related professionals, the main thrust of day-to-day casework still remains primarily that of the corroboration of identity of deceased persons. Historically, there are numerous accounts of famous persons being identified post mortem by the recognition of certain peculiarities within their dentitions.

Why is Dental Evidence So Good for Corroborating Identity?

Everyone has teeth. When they do not, it is common for people to wear prostheses. If they have never developed teeth, this is, in itself, so unusual as to be highly individualizing and may be only one manifestation of a syndrome with other, more obvious, physical signs.

The form of teeth and the detail of their arrangement in the dental arches provides a body of information that is probably unique to the individual. Even in identical twins, the slight variations in tooth form and position can enable the twins to be separated on the basis of their dentitions.

Add to the above criteria the almost infinite variation that can be added to the dentition by way of dental treatments (extractions, dental restorations, different use of restorative materials, the wearing of

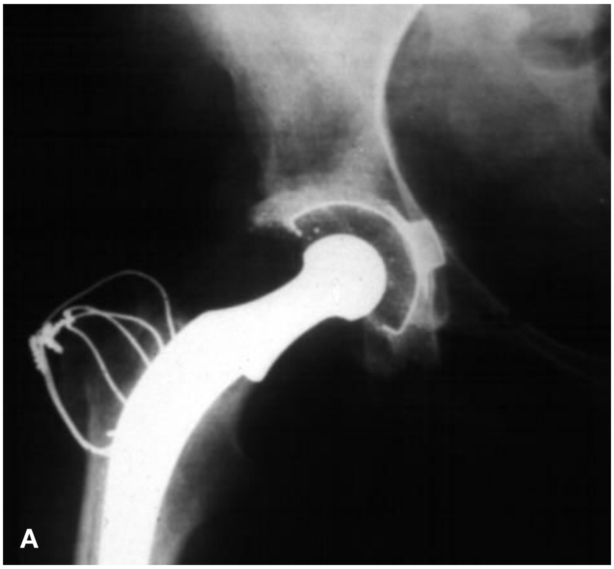
prostheses and orthodontic appliances) and the dentition is certainly as individual as a fingerprint.

One great advantage over fingerprint evidence for the corroboration of identity arises from the resistance of the oral structures to fire or putrefaction. Teeth and bone can be heated to temperatures approaching the melting point of skeletal mineral (>1600°C) without appreciable loss of microstructure or tertiary architecture. Furthermore, the teeth represent the only part of the skeleton that normally protrudes from the soft tissues of the body and is therefore available for examination and recording by photography or dental impressions before death.

This accessibility is the second advantage of dental evidence over fingerprints, particularly in societies where the taking and retention of fingerprints from persons is strictly limited to those with an existing criminal record. Similarly, much medical treatment returns the patient to full health with no residual indicators of a temporary disease permanently expressed; a sore throat is a transient event that leaves no scar and therefore nothing permanent and tangible with which postmortem comparisons can be made. Conversely, when some medical procedures, such as orthopedic surgery for a prosthetic hip replacement, are performed, the resulting records in all their forms are excellent for identification purposes **(Fig. 1)**. Unfortunately for identification purposes, such treatments are still not commonplace.

By comparison with such relatively uncommon surgical treatment, almost everyone in the developed world has had to attend a dentist for a check-up or treatment at some time in their life. The attendance at the dentist inevitably results in the production of records. These may be in the form of written clinical notes and odontograms (dental charts), plaster casts or radiographs ('X-rays') or even just an itemized bill for payment. All are potentially valuable but the standards of record keeping are highly variable. In the absence of comprehensive written or graphical records, the contribution of 'dental X-rays' to the process of corroborating identification can hardly be overstated.

Even when no treatment has been carried out in the form of restorations or fillings, X-rays can have been taken to check for the presence of tooth decay or to monitor the progress of periodontal diseases. In addition to the specific clinical need to take the X-ray, the resulting radiograph conveniently records many, normally inconsequential, features of the dentition and jaws. These characteristics include the size and shapes of tooth roots, dental restorations and the pulp chambers within teeth. The internal architecture of the bones of the jaws is often clearly revealed and may show the internal struts of bone, called trabeculae, the presence of healing tooth sockets post-extraction, the retained roots of teeth or the crypts containing developing teeth **(Fig. 2)**. In larger radiographs, such as those depicting panoramic views of the jaws (orthopantomograms (OPGs/or more strictly OPTs in dentists' jargon), the details may be less but there is a compensating increase in those structures imaged. These include the maxillary sinuses of the skeletal complex of the upper jaw. The paranasal sinuses are air-filled spaces or pneumatizations of the skull, all in communication with the upper respiratory tract,



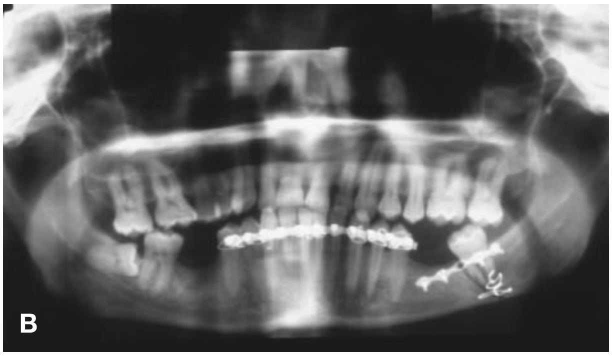


Figure 1 (A) Postoperative radiograph of a prosthetic hip joint. When such permanent evidence of medical treatment exists, it is excellent for corroborating the identity of a body where there is a presumption of putative identity. The type of prosthesis, the gauge of the wire, the number and nature of the turns, twists and loops in it, taken together are a signature of the surgeon and provide a unique image for comparative purposes. Such images are not commonplace. **(B)** This is the orodental equivalent of **(A).** A panoramic antemortem radiograph that images internal fixation of a reduced mandibular fracture. None of this metallic surgical material was seen at dissection, being completely internalized within the jaw after a long period of healing.

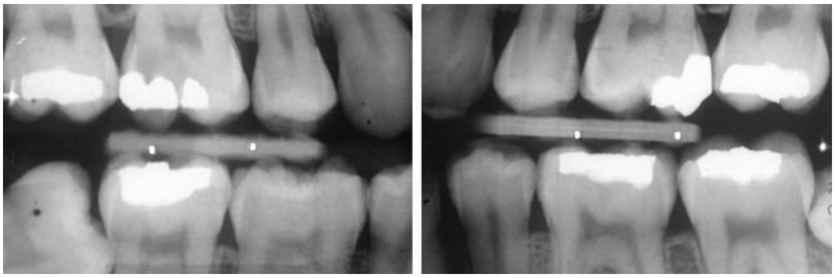


Figure 2 A pair of intraoral 'bitewing' radiographs. These are the most common views of the mouth, routinely taken by dentists to check for tooth decay or the presence or progress of degenerative disease in the supporting tissues of the teeth. Such projections also conveniently reveal the state of dental maturity in the young and summarize much of the restorative dental treatment experienced by the individual throughout life. In these particular radiographs the crowns of the, as yet unerupted, wisdom teeth can just be seen. The form of the other teeth, specifically the shapes of their roots in their sockets and the internal pulp chambers, is clearly seen. This patient has had eight dental amalgam restorations. Each restoration has a unique silhouette for any given geometric/radiographic projection.

which increase in size during early life but then stabilize in adulthood **(Fig. 3)**.

Similarly, when X-rays are taken of the whole head, other sinuses such as those buried within the facial bones of the forehead (the frontal sinuses) may be displayed. The size and shape of sinuses is very variable. Furthermore, they are often partially divided internally by walls or septa whose position is also stable and which are clearly visible on X-ray. The combination of these two features means that X-ray images of the sinuses can be excellent for comparison with postmortem findings, when the sinuses can be repeatedly X-rayed in slightly differing orientations until direct visual comparison of similarities and differences is valid. To achieve a valid match, and therefore a 'positive identification', all the features of interest must coincide simultaneously with all other features depicted on both radiographs. This requirement is an important check and a good internal standard to insure that a match of important features has not been unwittingly created. The process of comparison can be done by eye by directly overlaying antemortem and postmortem radiographs, or digitally by subtracting one image from the other when

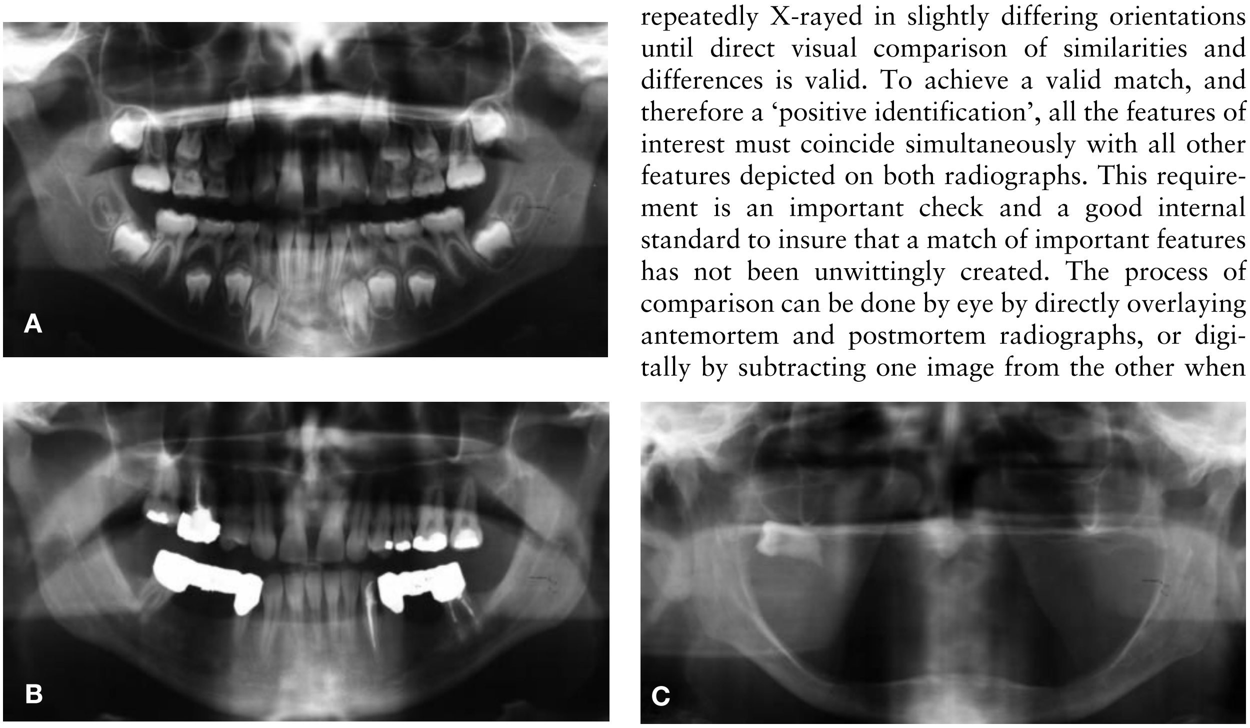


Figure 3 Three panoramic radiographs (orthopantomograms). **(A)** Radiograph from a child aged about 8 years. It shows a mixture of deciduous and permanent teeth erupted and functional within the oral cavity. Beneath them, within the jaws, there are many successional teeth developing in crypts. A knowledge of the chronology of tooth development and emergence sequence for the dentition allows age estimation to be undertaken with reasonable certainty during the first two decades of a person's life. **(B)** Radiograph from a middle-aged person aged 50. The patient has had multiple root canal treatments, during which finely tapered gutta percha points have been cemented into the root canals of several teeth. This style of treatment is now rather outdated. The combination of many dental restorations and root fillings has produced a unique pattern of features that is invaluable for identification purposes. **(C)** A radiograph from an elderly person aged 80. The natural teeth have all been extracted many years ago and full dentures have been worn ever since. This orthopantomogram reveals the presence of a single, completely formed tooth unerupted in the right maxilla. This tooth, previously undiscovered, has been buried within the jaws since the time of its initial formation over 60 years ago. The presence, form and position of the unerupted tooth, taken together with the myriad of other anatomical features depicted simultaneously, is unique.

all that remains is that which is different between the two views.

Sometimes the radiographs to be compared may look quite different to each other and yet yield evidence of a match of certain features. A good example might be if some teeth were present in the antemortem clinical X-ray image, yet missing in the corresponding postmortem view (perhaps lost perimortem as result of trauma or putrefaction). Supposing the feature to be compared was the shape and form of the sinuses, then, providing the tooth socket outlines still matched in size, shape and position, the missing teeth would not matter nor invalidate the match.

Another potential complication can arise from the passage of considerable time between the taking of the antemortem X-ray and the taking of corresponding views postmortem. This is particularly the case if the antemortem record has been taken when deciduous (baby) teeth are still present but the postmortem views are not taken until those deciduous teeth have been shed and replaced by permanent successors.

This serves as a good example of why an expert opinion is necessary to explain to scientific lay people within the legal system that, although the pictures look different, they are really of the same person; just taken at different stages of their life, which is explicable if the chronology of dental development is understood **(Fig. 4)**.

When No Antemortem Records Exist

So far, the assumption has been made that antemortem records exist and, once located, can be used for comparison with postmortem findings. However, records are not always traceable, or retained. In about a third of cases (32%) in Victoria, Australia (a state with high immigration), no records are available for the odontologist. In such cases, what inferences can be made?

Teeth develop in a predictable sequence, which enables age to be estimated during their formation with reasonable accuracy. This period extends from midfetal life until the attainment of physical adulthood in the early part of the third decade. In the absence of birth records, the eruption of second permanent molars formed the basis upon which children were deemed old enough to work in factories during the industrial revolution in Britain in the early 1800s: a widespread forensic application of dental knowledge.

During adulthood, the teeth and their supporting tissues undergo progressive changes. These can be greatly influenced by diet, habits, customs and lifestyle. In this context the front-line work of the forensic odontologist differs little from any discipline concerned with reconstructing past events from incomplete evidence and the interpretation of their significance. In archeology, paleontology and biological anthropology, there is a forced reliance upon inference and comparison with accumulated knowledge rather than a direct 'concrete' comparison between records and findings. The results of such inferences need to be carefully presented to avoid the impression of unwarranted certainty. Comprehensive accounts of determination of race, gender and age from the orofacial structures have been published **(Fig. 5)**.

For example, in the 1960s Gustafson consolidated many of the age-related changes of the teeth into a unifying series of age predictive regression equations. He scored the progressive infilling of the dental pulp chambers by secondary dentine, attrition caused by tooth-to-tooth contact under the influence of the diet, compensatory cementum deposition, migration of soft-tissue gingival attachments, and increasing translucency of dentine due to increased peritubular (strictly, intratubular) dentine deposition. Several other researchers have critically revisited and revised all or part of Gustafson's method, introducing minor modifications from time to time **(Fig. 6)**.

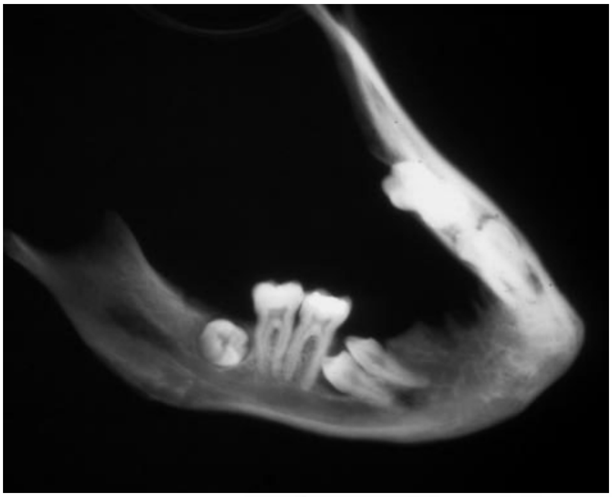
In essence, Gustafson's method was probably successful at that time because of the homogeneity of his Swedish sample in terms of genes, social and lifestyle factors and diet. The methods are of much less use in countries such as Australia, where there has been recent, large-scale immigration from many different places by peoples of different ethnicity, each with their own legacy of past diet, habits and customs.

Of all the age-related factors described above, the increasing translucency of dentine with age is probably the most physiologically related and hence the least affected by environmental influence.

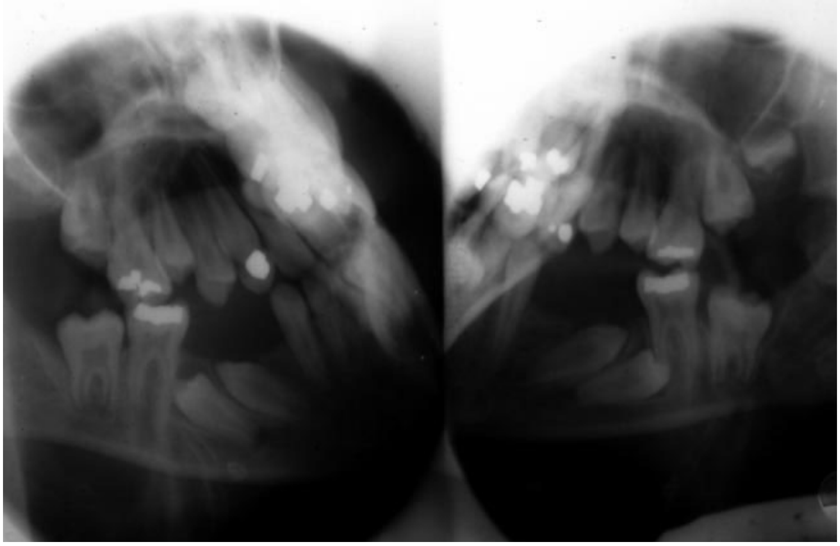
In common with other products of complex epithelial/mesenchymal interactions, such as fingerprints, the teeth too reveal evidence of heritability in some of their features. This results in ethnic differences in the dentition, which can be of forensic importance. Anthropologists who study populations rather than individuals for evidence of individual identity tend to use large numbers of physical measurements to investigate racial traits. This metric approach has spawned many studies on tooth crown dimensions, often from plaster casts obtained in the field or directly from living individuals over a period of time. The results are of limited forensic value.

Interactions with Other Professionals

For the majority of cases it is expected that the odontologist will form part of a team of investigators. Such teams may comprise a pathologist, police



A



B

Figure 4 A homicide case where the mandible is the only part of the skull to have been retrieved after prolonged immersion of the body. **(A)** Lateral oblique view of the recovered skeletalized mandible shows a very unusual developmental malpositioning of the mandibular premolars. The malformation was bilateral. **(B)** Two lateral oblique views (left and right) of the same person taken during life by the school dental service. The misplaced premolars can clearly be seen on both sides of the jaw. More teeth are present in these earlier clinical radiographs but the difference between **(A)** and **(B)** is explicable when the intervening period is taken into account.

investigators and scene of crime personnel, mortuary staff, photographers, sculptors, artists and other experts, such as entomologists, molecular biologists or accident investigators. In cases where living victims of assault have to be examined and treated, the composition of the team will differ but is likely to include hospital doctors, nurses and the staff of the local government social services department.

Whatever the composition of the group, clear lines of responsibility and reporting need to be understood by all involved. The coordinator/facilitator/leader (usually the most senior police officer) has to have a clear overview of the evidential requirements for the entire investigation and to be aware of the sometimes conflicting requirements of the investigators. For example, it would be a catastrophe for the forensic odontologist to adopt the policy of slavishly resecting the jaws of deceased victims, in order to make examination of them easier, if the only available evidence that would have corroborated identity required the facial skeleton to be kept intact. The best

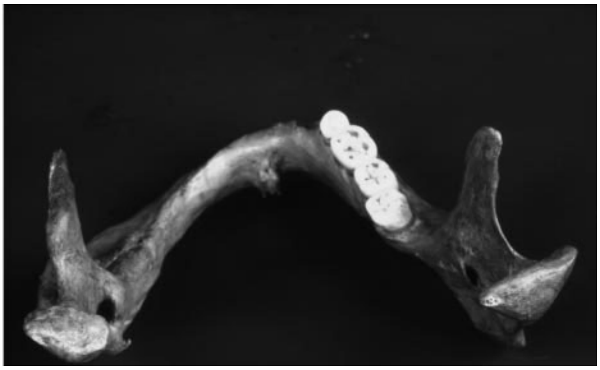


Figure 5 An archeological specimen. This mandible was found on the Chatham islands east of New Zealand. The specimen carries a history of a fracture that at one time split the body of the mandible lengthways. Teeth were lost in the fracture line, either at the time of the injury or during subsequent infection before eventual healing. The bony fragments are misunited because the wound was never reduced by a doctor. As a consequence of the loss of teeth the man was obliged to chew on one side of the jaw only. Those teeth show heavy attrition. The jaw joint on one side has suffered from osteoarthritic degeneration. From the above observations and inferences it would be highly unlikely that this person would have lived in the last 100 years. The find is therefore of little interest in a coronial context.

Figure 6 Age changes to the teeth. This longitudinal section through a tooth has been X-rayed to produce a microradiograph. The tooth is from a middle-aged person. It shows attrition through the enamel to dentine, and reactionary changes in the underlying dentine. The cervical margin (neck) of the tooth reveals cervical abrasion after gingival recession has taken place. To compensate for the loss of enamel from the crown due to wear, more cementum (a bone-like substance) has been deposited at the tip of the root. These and other changes were those quantified by Gustafson.

investigations are undertaken by groups that comprise experts who understand each others' skills, perhaps because they have worked together on numerous occasions in the past, but in all cases it is important for each expert to know what is required of him or her. A case conference at the outset of a multidisciplinary investigation quickly repays the time and energy expended by assuring efficiencies and precision later on. Sometimes what was initially seen as an apparently uncomplicated 'deceased person identification', undertaken on behalf of the coroner, can later be discovered really to be an event requiring a full homicide investigation. In many jurisdictions the evidential requirements for different courts differ in their rigor. In the UK system of justice, and many of its derivatives, the coroner requires proof of identification only 'on the balance of probability', whereas a higher court which has to try a charge of murder requires identification to be established more stringently 'beyond reasonable doubt'. It is therefore important for the forensic odontologists to anticipate where each case may lead and to conduct their investigation with the final forum for their deliberations clearly in mind.

Bitemarks, Bruising and Other Injuries to Skin

Forensic odontologists are frequently called to look at people, both dead and alive, or objects which may retain 'tool marks' left by an interaction with the teeth of a person or an animal or animals (or something which closely resembles such marks). The circumstances range from excellent, clearly identifiable, bitemarks left by an individual in a dimensionally stable substrate such as hard cheese or chocolate, to the badly mauled, barely recognizable remains of a deceased person killed by a large carnivore or carnivores **(Fig. 7)**. The situation can be even more difficult if the case is one involving postmortem animal predation of the corpse, where scattering of the remains and superimposed decomposition add complexity. Where a person has been mauled to death and the animal which is thought to have caused the injuries can be caught or killed, it is relatively simple, by the use of emetics or dissection, to prove the involvement of the animal. In instances where humans bite other persons, it is uncommon for large pieces of tissue to be removed. The marks remaining from the bite may range from an excellent silhouette of the dental arches and individual teeth to something very diffuse and indistinct; the latter case is unfortunately the far more common circumstance.

The poor fidelity of human skin as a recording medium, its deformability and elasticity, its capacity to heal and the curved nature of the external surface of the body, added to the complex form of the dental arches and the many possible movements of the jaws during biting, combine to make scientific analysis of bite mark injuries very problematic. While there are scientific techniques, such as holography, that can accurately record curved three-dimensional shapes for analysis, they are practically useless for 'real-

Figure 7 Shark bites. **(A)** A life jacket recovered after a yacht was lost. The jacket carries serrated bitemarks left by a large shark. **(B)** (see color plate 34) A postmortem shark bite on the remains of someone who had drowned. The arcing lacerations with regular spacing are quite typical of a medium-sized shark bite.

world' bitemark investigations, in which the distressed, living victim often has to be examined in the restricted environment of a hospital accident department or, worse, in a police station.

Where some topography in the form of indentations, cuts, abrasions or exudates can be found on the victim, the recording of bitemarks is customarily done by means of dental impression materials and swabs for biological analyses. Impressions taken from such injuries are used to make rigid models in plaster or dental stone. These models can then be used for comparison with the dentition or models of the dentition of the suspected assailant (or victim if a self-inflicted injury is possible). The use of such models is usually augmented by photographs.

Where the injuries are less severe and/or more diffuse, photography may be the only practical method for documenting the injuries. Light outside the spectrum visible to the human eye can be used to produce images that reveal bruising, which may be undetectable using visible light. Exactly what histological or chemical changes to the skin enable these images of old and/or 'latent' bruising to be made still have to be determined.

Without the foundations of fundamental scientific knowledge about the biology and chemistry of bruising, and the ability to model accurately the actions of the teeth and jaws interacting with the soft tissues of the body in a realistic, dynamic way, the presentation of bitemark evidence to the courts will remain problematic. In Australia, bitemark evidence has become so discredited in the wake of two prominent cases that drew heavily upon the opinion of foreign experts, and in which it has since been judged that some of the experts may have overinterpreted their observations, that in a recent protocol set down for the taking of 'intimate samples' bitemark evidence was considered to be of little value. This unfortunate extrapolation overlooks the great *exculpatory* value of bitemarks to an investigation and the justice system.

Other Skills

Implied in the previous section is a requirement for the forensic odontologist to have a comprehensive knowledge of comparative dental anatomy. At its most obvious, there is a common need to be able to sort animal remains from those of human beings. While some anthropologists and archeologists are experts in this, very few understand the dentition as well as their dental colleagues, particularly where only particulate remains or fragments of skeletal material are found and a study of microstructure is required. This is important in forensic cases where contamination of foodstuffs is alleged by a consumer.

In those countries where people are still killed by wild animals it is important to be able to recognize the characteristic pattern of injuries inflicted by the local predators.

Summary

At this time, forensic odontology remains one of the most reliable, cheapest and quickest means of corroborating human identity **(Fig. 8)**. By its very tangible nature it is easily presented to, and understood by, juries. In years to come, many of these identification tasks are likely to be answered by advances in molecular biology. DNA evidence will become more reliable and, more importantly, less controversial in court. Age-at-death determination from predictable and progressive changes to the organic matrices of hard tissues is a technique that has not yet become widespread, despite important pioneering studies showing its feasibility. However, while much of what is done today may become redundant, other aspects of the work of the forensic odontologist are in the ascendancy. Forensic sculpting upon remnant skull evidence is growing in popularity as a technique

Figure 8 An anatomical replica in methylmethacrylate, made using dental techniques for use in a court room demonstration by the pathologist years after the interment or cremation of the remains. Three-dimensional replicas are much more successful than two-dimensional photographs for displaying osseous injuries, such as hatchet wounds or bullet trajectories. Courtesy of R W Taylor.

of last resort, and the scientific basis upon which reconstructions (approximations) are based is becoming more rigorous. There is a large degree of overlap of these traditional techniques with newer computer-based techniques replacing 'Photofit' and 'Identikit', together with psychological research into facial recognition, with all its cultural and ethnic overtones. The very precise anatomical modeling of human skeletal remains, as aides mémoires for the pathologist and as court exhibits long after the interment or cremation of the deceased, has its basis in dental materials and technology. These skills will always be in demand.

However, to return to our roots for a moment; the forensic odontologist is a person with dental qualifications and expertise who also has considerable experience of the law and courts and issues relating to the ethics and conduct of the dental profession at large. It is commonplace for police faced with allegations of professional misconduct or charges of physical or sexual assault at the hands of a dentist in a clinical setting to consult the forensic odontologist. He or she is seen by investigators as a professional peer, who understands what is deemed to be acceptable contemporary clinical behaviour in dentistry before considering whether a prima facie case exists that could warrant prosecution.

The investigation of illegal, unregistered practitioners, and the premises from which they operate, rightly involves the forensic odontologist, who is able to advise on health and safety aspects for the investigators and the status of prescription medicines recovered at the scene. These are but two examples, but other important issues pertaining to consent to treatment, confidentiality, access to records, malpractice, litigation, risk minimization and evidence-based practice are subjects to which much more prominence has been devoted in newer medical school curricula where departments of legal medicine have been newly established. It is highly likely that those developments will soon be mirrored in dentistry, and the forensic odontologist will rightly be in the vanguard of such advances within our society.

See also: **Crime-scene Investigation and Examination**: Fingerprints. **Fingerprints (Dactyloscopy)**: Identification and Classification; Standards of Proof.

Further Reading

Ciapparelli L (1992) The chronology of dental development and age assessment. In: Clark DH (ed.) *Practical Forensic Odontology*, pp. 22–42. London, UK: Wright.

Clement JG (1998) Dental identification. In: Clement JG and Ranson DL (eds) *Craniofacial Identification in Forensic Medicine*, pp. 63–81. London: Arnold.

Clement JG (1993) Forensic dentistry. In: Freckelton I and Selby H (eds) *Expert Evidence: Practice and Advocacy*, vol. 1. Australia: Law Book Company.

Clement JG, Officer RA and Sutherland RD (1995) The use of digital image processing in forensic odontology: when does 'image enhancement' become 'tampering with evidence'? In: Jacob B and Bonte W (eds) *Advances in Forensic Sciences*, Vol. 7: Forensic Odontology and Anthropology, pp. 149–153. Berlin: Köster.

Demirjian A and Goldstein H (1976) New systems for dental maturity based on seven and four teeth. *Annals of Human Biology* 3:411–421.

Gustafson G (1966) *Forensic Odontology*. London: Staples Press.

Hillson S (1996) *Dental Anthropology*. Cambridge: Cambridge University Press.

Kraus BS and Jordan RE (1965) *The Human Dentition Before Birth*. Philadelphia: Lea and Febiger.

Kvall SI (1995) *Age Related Changes in Teeth. A Microscopic and Radiographic Investigation of the Human Permanent Dentition*. PhD thesis, Oslo: University of Norway.

Logan WHG and Kronfeld R (1933) Development of the human jaws and surrounding structures from birth until the age of fifteen years. *Journal of the American Dental Association* **20**:379–427.

Moorrees CFA, Fanning EA and Hunt EE (1963) Age variation of formation stages for ten permanent teeth. *Journal of Dental Research* **42**:1490–1502.

Ogino T, Ogino N and Nagy B (1985) Application of aspartic acid racemisation to forensic odontology: postmortem designation of age at death. *Forensic Science International* **29**:259–267.

Ritz S, Schultz HW and Schwarzer B (1990) The extent of aspartic acid racemization in dentin: a possible method

for a more accurate determination of age at death? *Zietschrift für Rechtmedizin (Journal of Legal Medicine)* **103**(6):457–462.
Saunders E (1837) *The Teeth as a Test of Age, Considered with Reference to the Factory Children: Addressed to the Members of Both Houses of Parliament*. London: H Renshaw.
Smith BC, Holland MM, Sweet DL and Dizinno JA (1995) DNA and forensic odontology. In: Bowers MC and Bell GL (eds) *Manual of Forensic Odontology*, pp. 283–298. Colorado Springs: American Academy of Forensic Sciences.
Sognnaes RF, Rawson RD, Gratt BM and Nguyen NBT (1982) Computer comparison of bitemark patterns in identical twins. *Journal of the American Dental Association* **105**:449–451.

OFFENDER SIGNATURE

B E Turvey, Knowledge Solutions LLC, Watsonville, CA, USA

Copyright © 2000 Academic Press

doi:10.1006/rwvi.2000.0752

Introduction

In the discipline of criminal profiling, signature is a term used to describe the behaviors committed by an offender that serve the person's psychological and emotional needs. It is through analyzing and interpreting a particular offender's signature behavior in a crime scene, in combination with other elements such as modus operandi (MO) and victimology, that criminal profilers behaviorally link cases and develop an understanding of an offender's motives. The successful interpretation of fantasy and motivation through offender signature involves an appreciation of both psychology and physical evidence and can be found most commonly in the investigative use of offender motivational typologies. It is commonly held that an individual offender's signature aspect tends to change less over time than the MO, making it a more reliable element for behavioral case linkage. However, signature behavior may also evolve over time with the development of an offender's fantasy. To recognize signature behaviors separately from MO behaviors and use them to link cases and gain insight into offender motivation, it is important to understand the basis of signature in human development and the elements that can affect it.

Definitions

Offender 'signature' has been defined as any behavior committed by an offender beyond that which is necessary to successfully complete a crime. There are two separate but dependent parts to the concept of signature. First, there is the general 'signature aspect' of a crime. This represents the emotional or psychological theme that the offender satisfies by engaging in certain constellations of crime-related behaviors. Second, there are the actions themselves, referred to as 'signature behaviors'. They include crime-related actions that are not necessary to commit the crime successfully. Signature aspect is manifested or evidenced by particular constellations of signature behaviors.

Signature is a term that was first used by the American criminal profiler John Douglas, a supervisory special agent with the Federal Bureau of Investigation who eventually came to direct their Behavioral Sciences Unit. He developed the term to help those investigators involved in criminal profiling distinguish offender behaviors that suggest psychological needs and themes from offender behaviors that are a part of the MO. Despite the development of the term and its adoption by the law enforcement and legal communities, there is still confusion over its use on a case-by-case basis.

The confusion is not only the result of the dearth of peer-reviewed research on the subject. Each offender has his or her own varied knowledge and skill which influences the MO behaviors. Each offender expresses psychological needs during crime-related

activities differently. Each individual crime scene contains a multitude of variables that can influence offender actions. When the offender involves a victim, each one brings his or her own set of reactions to offender behavior, which the offenders in turn have a reaction to. Given all of these things, signature behaviors are not always readily apparent, even to the most competent criminal profilers. In the case of one offender, for example, the act of covering a victim's face with her own shirt during a rape may evidence a psychological need, facilitating a fantasy that the victim is another person. This would be a signature behavior. In the case of another offender, the act of covering a victim's face with her own shirt during a rape may evidence a desire to keep the victim from seeing his face and identifying him at a later time. This behavior would then be considered a part of the offender's modus operandi.

Psychology of Offender Signature

Not all offenders are alike. Similar behaviors committed under similar circumstances by different offenders will not necessarily be for identical or even similar reasons, and this is related directly to influences on normal human development.

According to the American psychologist John Money, the reason for this distinctiveness is that offenders have in their mind's eye a pattern of specific behaviors, and subsequent associated feelings, that he refers to as a 'love map'. Love map is a term Money developed to describe an idealized scene, person and/or program of activities that satisfy the particular emotional and psychological needs of an offender. Love maps, needs or fantasies develop in all people as a part of the natural process of human development and can subsequently be affected by both biological and environmental factors. Money theorizes that criminal behaviors result when the human developmental process is derailed, and a person is able to make pleasurable associations with violent or otherwise criminal activity. These associations, varied and evolving over time, amount to a behavioral distinctiveness in the way that an individual offender seeks to satisfy emotional or psychological needs during the commission of a crime such as rape, homicide, arson and other similar or serial offenses.

The specific etiology of offender signature has been described as an offender's fantasies that are progressive in nature and contribute to thoughts of committing violent or predatory behavior. As an offender's fantasy behavior develops over time, so does the need to live out those fantasies. When a violent or predatory fantasy is subsequently acted out, the act itself fuels the fantasy in the mind of the offender and causes it to evolve. The process is complementary and can facilitate the evolution of fantasy, signature behaviors and signature aspect over time.

Signature Behavior

The distinctiveness of a particular offender's needs is manifested by his or her particular pattern of signature behavior. Signature behaviors, therefore, are best understood as a reflection of the underlying personality, lifestyle and developmental experiences of an offender. With the potential of being fairly distinct to a particular offender, depending upon the influences on crime-related behavior and the available evidence of that behavior, a specific pattern of signature behaviors and the needs that they represent may be used investigatively to distinguish between two offenders.

Examples of individual signature behaviors include, but are not limited to:

- special order/type of sexual activity;
- special type of ligature or binding;
- infliction of special/specific types of injuries;
- display of the body for shock value;
- torture and/or mutilation of the victim;
- engagement in ritualistic/repeated behavior;
- domination;
- manipulation;
- control;
- excessive vulgarity and/or abusive language;
- scripting the victim;
- excessive brutality.

Signature behaviors are the manifestation of offender needs within a crime scene. They are evidenced in the interaction between the victim, the offender and the crime scene itself. A convergence of offender signature behaviors is used to infer the signature aspect or motivational theme of the crime.

Signature Aspect and Motivational Typologies

Signature aspects are the offender's motivational themes, or psychological needs, suggested by convergences of signature behaviors. They are thought to be relatively stable over time within offenders, although they are susceptible to the influences of fantasy development.

In order to understand the motivations of violent, predatory offenders, Nicholas Groth, an American clinical psychologist who worked with both victims and offender populations, published a study of over 500 rapists in 1979. In that study, he found that rape, like other crimes that satisfy emotional needs, is

complex and multidetermined; that is to say, the act of rape itself serves a number of psychological needs and purposes for the offender. From Groth's studies and the work of others, a rapist motivational typology has been developed that places offender behavior into one of five typologies (which are examples of signature aspects, in criminal profiling terminology):

- *Power reassurance*: offender behaviors that suggest an underlying lack of confidence and inadequacy, or belief that the offense is consensual, expressed through minimal force and low confidence (i.e. fondling and foreplay behavior, apologies, compliments, minimal victim injury, weapon present for control only, etc.).
- *Power assertive*: offender behaviors that suggest an underlying lack of confidence and inadequacy, expressed through a need for control, mastery and humiliation of the victim, while demonstrating authority (i.e. use of extreme profanity, macho attitude, verbally commanding/threatening, limited fondling and foreplay behavior, ripping of victim clothing, weapon used to inflict corrective injury).
- *Anger retaliatory*: offender behaviors that suggest a great deal of rage, either towards a specific person, group, institution or a symbol of either (i.e. use of brutal force, high amount of potentially lethal victim damage, hostile language, weapons of opportunity, lack of planning, etc.).
- *Sadistic*: offender behaviors that suggest that the offender gets sexual gratification from victim pain and suffering (i.e. physical, mental, or sexual torture of a living victim, time spent enjoying the attack, etc.).
- *Opportunistic:* offender behaviors that suggest an offender who is out to satisfy immediate sexual impulses (i.e. rape committed during another offense, victim of opportunity, lack of planning, much physical evidence left behind at the scene, etc.).

Although it does not provide a dynamic scale that measures a rapist over time, it is an excellent way of assessing them at a particular moment. The typologies are not to be confused with profiles, nor by any means should they be considered an exclusive list of signature aspects. They provide a psychological snapshot of a rapist during a single instance from which some inferences can be made. There is also no bright yellow line between the typologies, meaning that a single offender can evidence aspects from more than one. The typologies actually classify the needs satisfied by offender behavioral patterns at one point in time, rather than define the offenders themselves.

As with motivational typologies, signature aspects are not always exclusive. A single offender may evidence multiple signature aspects during the same offense and separate signature aspects during different types of offenses. Additionally, a single behavior cannot be taken out of context to suggest an overall pattern. Patterns only become evident upon the convergence of multiple behaviors in a context that is known and understood by the profiler.

One example would include an offender who, within the same attack over a period of a few hours, tortures a prostitute with a pair of pliers and yells at them to scream so that he can hear it, and then engages in caressing and fondling behavior while apologizing. The context is the rape and possible abduction of a stranger. The evidence of torture would suggest a sadistic signature aspect, while the caressing, fondling and apologizing would suggest a reassurance-oriented signature aspect.

Another example would include an offender who robs a convenience store on one day, shooting all of the patrons inside at least six times each in full view of the security camera, but who robs a bank a few days later, wearing a disguise, using a note, without firing a single shot. The amount of overkill and lack of concern for later identification in the convenience store example begins to suggest an anger aspect. This is a marked contrast from the lack of psychologically revealing behaviors evidenced in the profit-motivated bank robbery.

Recognizing Offender Signature

As mentioned, the term signature and the underlying psychological concepts have been developed specifically to separate them from an offender's MO. Because MO behaviors can change over time, the tendency of law enforcement to rely solely upon an offender's MO for investigative strategy and case linkage can lead to what has been termed 'linkage blindness.' Linkage blindness has been defined as the failure to recognize a pattern that links one crime with another crime in a series of cases.

Specifically, there are three very important factors that can act individually or in concert to cause linkage blindness:

- The tendency for law enforcement to rely solely on MO behaviors, such as victim type, weapon selection and location type, as a basis for case linkage.
- The possibility that one predatory offender is operating in or near the same general area as another, confusing law enforcement efforts.
- Interpersonal or interagency conflicts, which can lead to communication breakdowns and a lack of information sharing.

The ability to read the behavioral evidence in a crime scene and recognize offender signature has been

shown to be extremely beneficial to the investigative process, and there are many examples of the successful investigative use of offender signature in the published literature.

There are important limitations on the concept of signature that must be understood. It is true that many serial or predatory offenders have a need to engage in personal expressions that are unique only to their individually formed personality. However, despite this behavioral distinctiveness, a result of the many different variables affecting the human developmental process, it is not truly appropriate to state that two crime scenes related by signature alone are psychologically 'identical'. The terms 'identical' and 'match' have very absolute connotations in the forensic sciences and can be misleading to those who do not fully understand the concept and psychology of offender signature.

One of the primary reasons for the lack of absolute certainty in interpreting signature behaviors as unique to a specific offender is the subjectivity of the interpretation itself. While offenders may be psychologically distinct, profilers cannot see through the eyes of an offender with perfect, objective clarity. They can show the most likely perspectives and needs of the offender by demonstrating a strong convergence of the physical and behavioral evidence, but they cannot go so far as to call it a 'psychological fingerprint'.

In addition, there are many variables to consider when interpreting signature behaviors. These must be factored into any complete analysis. It is important to understand that it may not always be possible to link related cases with signature, because an offender may not always leave his or her signature behind. When making inferences about offender signature, profilers may not have all of the facts in the case, or may be operating with flawed investigative assumptions, and must consider the following:

- Whether the amount of behavioral evidence is competent and sufficient to make an interpretation of offender signature (i.e. was a competent crime scene reconstruction performed, are the forensic protocols that the reconstruction is based on competent, was evidence of wound patterns lost due to bodily decay, etc.).
- Whether the amount of behavioral evidence is fully representative of the offender's needs (i.e. is there evidence of interruption during the crime, did the offender have the time to do all of the things that were felt to be important, etc.).
- Whether the behavioral evidence suggests a signature that is part of the escalation or the evolution in an offender's fantasy continuum, or whether the signature appears to be relatively fixed over time.

Given these pitfalls, it is most appropriate to explain the nature of an offender's patterns of signature behavior, subsequent case linkages, signature aspects or other inferences in terms of an appropriate level of confidence addressing the considerations mentioned. Understanding offender signature is an important investigative tool and an important part of understanding offender fantasy and motivation. However, until such time as a consistent concept of offender signature has been more thoroughly documented in the literature and is more universally understood in terms of practical application, a level of caution and care is requisite.

See also: **Crime-scene Investigation and Examination**: Scene Analysis and Reconstruction. **Criminal Profiling**. **Modus Operandi**.

Further Reading

Burgess A and Hazelwood R (eds) (1995) *Practical Aspects of Rape Investigation: A Multidisciplinary Approach*. New York: CRC Press.

Burgess A, Douglas J and Ressler R (1988) *Sexual Homicide: Patterns and Motives*. New York: Lexington Books.

Burgess A, Burgess A, Douglas J and Ressler R (1997) *Crime Classification Manual*. San Francisco: Jossey-Bass, Inc.

Egger S (1997) *The Killers Among Us*. Englewood Cliffs, NJ: Prentice Hall.

Geberth V (1996) *Practical Homicide Investigation*, 3rd edn. New York: CRC Press.

Groth AN (1979) *Men Who Rape: The Psychology of the Offender*. New York: Plenum Press.

Holmes R and Holmes S (1996) *Profiling Violent Crimes: An Investigative Tool*, 2nd edn. Thousand Oaks, CA: Sage.

Money J (1989) *Lovemaps: Clinical Concepts of Sexual/Erotic Health and Pathology, Paraphilia, and Gender Transposition in Childhood, Adolescence, and Maturity*. New York: Prometheus.

Turvey B (1999) *Criminal Profiling: An Introduction to Behavioral Evidence Analysis*. London: Academic Press.

P

Packaging *see* **Crime-scene Investigation and Examination**: Packaging.

PAINTS AND COATINGS

Commercial, Domestic and Automotive

L Brun-Conti, Michigan Department of State Police, East Lansing, MI, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0581

Paint as Forensic Evidence

The forensic chemist will receive paint as evidence in a number of different forms. Automotive paint, as chips or smears, comes into the laboratory from hit-and-run accidents when the offending vehicle strikes a person or another vehicle. If a vehicle hits a person, paint chips may be found in the victim's clothing or on the person. Structural paint may be found on breaking and entering tools (prybars, screwdrivers). A double transfer may occur in any of the preceding cases. A double transfer is where the paint from the suspect item (vehicle or tool) is transferred on to the victim item, and paint from the victim item is transferred on to the suspect item. For example, paint from the tool the suspect used to break into a home is transferred on to the structure, as well as the paint from the structure being transferred on to the tool. Paint from a vandalism may come in the form of sprayed graffiti from the outside or inside of a building. In general, anything that is coated with a paint or protective coating has the potential to contribute trace amounts of the coating in a transfer. As Locard's principle states, when two objects come into contact with one another, each of the objects may leave particles of one on the other.

Paint (generally a liquid but in some cases a powder, see Application Techniques) is composed of three basic elements: the vehicle, the pigments and extenders, and the solvent. The following sections will describe these elements of the paint.

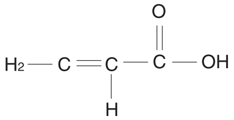
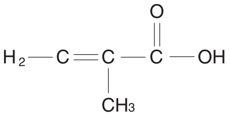
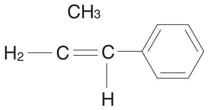
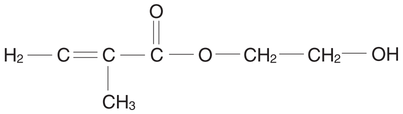
Vehicle or Polymeric Portion of Paint

Types of polymer

The vehicle portion of the paint gives the final film properties, such as resistance to acids, alkalis, solvents, water or any other material that would mar the finish from the outside. Also, the vehicle formulation will effect adhesion, hardness, flexibility, abrasion resistance and gloss. Some of the following vehicle polymers may be used in combination with another so that the desired properties of the final film are achieved. The vehicle polymer resins include acrylic, alkyd (or polyester) epoxy, urethane, amino, vinyl, phenolic and cellulose. The polymer system used will determine whether the final film is a thermoset or nonconvertible film, which, once in a dry film, cannot be converted back into a liquid state, or a thermoplastic or convertible film, which can be converted back into a liquid state by the addition of solvent.

Acrylic resins Acrylic resins are polymers made up of functional monomers (containing sites for crosslinking) and nonfunctional monomers (which do not contain additional sites for crosslinking). These monomers are polymerized into polymer sections of a controlled size. Thermosetting acrylic resins have active sites on the polymer to facilitate crosslinking. The type of monomers used in the formulation of the polymer will depend upon the properties desirable in the final finish. **Table 1** lists are four examples of common monomers used in acrylic resins and the properties

Table 1 Structure and properties of some common monomers used in acrylic resins

Monomer	*Structure*	*Properties*
Acrylic acid		This carboxyl functional monomer aids in the reaction of hydroxyls on the acrylic polymer and crosslinking resin to promote cure. Too high a level can cause problems with brittle films, gloss loss from humidity and poor exposure durability. Too low a level yields poor hardness or cure, poor humidity resistance and poor outdoor durability
Methacrylic acid		The same comments apply to methacrylic acid as to acrylic acid. Methacrylic acid had a slightly higher Tg (higher potential to become hard and brittle)
Styrene		Provides hardness and chemical resistance but poor outdoor exposure durability
2-Hydroxy ethyl acrylate		This monomer is important as it yields all durability and hardness properties. This is because the hydroxyl is where the acrylic and crosslinking resin are joined to form the crosslinked network

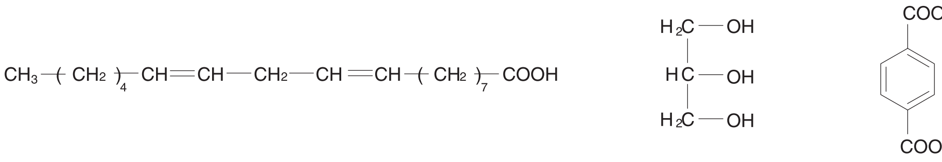
they impart to a film. Polymer chemists must formulate the acrylic resin using just enough, but not too much of the monomers available to them, keeping in mind the properties which they desire and which monomers will impart those qualities to the film.

Alkyd and polyester resins Polyester resins have three basic components which polymerize to form the resin: fatty acids, polyols and dibasic acids (**Fig. 1**).

The fatty acid and the polyol form an ester linkage. The dibasic acid is then added to the mix. The molecular weight of the polyester resin will depend upon the quantity of the dibasic acid and additional polyol. The polyol used is generally glycerol, but ethylene glycol (two –OHs), pentaerythritol (four –OHs) or sorbitol (six –OHs) may be used. The dibasic acids used are usually phthalic anhydride, isophthalic acid and teraphthalic acid. The fatty acids and glycerol are occasionally replaced by oils (three moles of a fatty acid plus one mole of glycerol is a triglyceride or oil) and the resin is then known as an alkyd. When using oils in an alkyd, the alkyd is defined by amount of oil: a long oil alkyd contains $>55\%$ oil, a medium oil alkyd contains 31–55% oil, and a short oil alkyd contains $<30\%$ oil. The oils are made up of two or three different fatty acids which will have varying amounts of unsaturation. In essence, the more unsaturation or double bonds, the more readily the resin will dry. This becomes important when considering oxidation curing ability of a film (see Film formation).

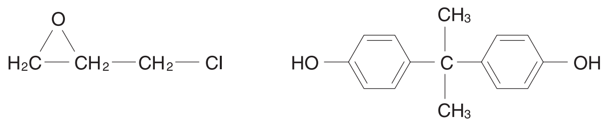
Epoxy resins Epoxy resins are made by polymerizing an oxirane ring-containing compound, such as epichlorohydrin, with bisphenol A (**Fig. 2**). The reactivity of the oxirane ring structure depends upon whether the oxirane is internal or external. In most epoxy resins the oxirane ring is external.

Urethane resins Urethane resins are basically the result of a reaction between an isocyanate and an active proton (**Fig. 3**). The reactive proton can be (in order of reactivity) alkyl-NH_2, aromatic, R-NH_2, primary –OH, secondary –OH, tertiary –OH, aromatic –OH, carboxylic –OH, primary amide. The isocyanate may come in an intermediate form as a diisocyanate. When the isocyanate is added to the active proton-containing resin, the crosslinking begins immediately and therefore must be used within a limited period of time. This 'two-package' system may be cumbersome to use. To make urethanes more convenient to use a few different 'one-package' urethane systems were



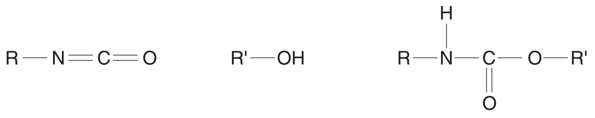
Linoleic acid (fatty acid) Glycerol (polyol) Teraphthalic acid (dibasic acid)

Figure 1



Epichlorohydrin Bisphenol A

Figure 2



Isocyanate Active proton Urethane

Figure 3

developed. One type is a moisture cure system where the polyol diisocyanate adducts cure primarily through the reaction of the terminal N=C=O with atmospheric moisture (**Fig. 4**). The resulting $-NH_2$ acts as a catalyst to continue the reaction.

Another one-package urethane system is a blocked adduct urethane coating. In this system the polyol is mixed in the same container with a blocked isocyanate monomer. Once the coating is applied and baked, the blocking component is split and the isocyanate group is free to react with the polyol.

Vinyl resins Vinyl resins are made of the ethylene derivatives illustrated in **Fig. 5**. The vinyl resins are polymerized using addition polymerization (see Polymerization). Polyvinyl chloride and polyvinyl acetate are prevalent in latex architectural paints.

Phenolic resins Phenolic polymers are based on the reaction of phenol with formaldehyde (**Fig. 6**). When an acid catalyst is used for polymerization, the polymers become a two-dimensional polymer called a 'Novolak'. Novolaks produce hard, nonheat-reactive, chemically resistant coatings. An alkaline condensation of the phenol and formaldehyde monomers forms a resin called 'bakelite'. During this process, the first stage (stage A) produces a monomer with three methylol groups ($-CH_2OH$), the second stage (stage B) produces a polymer with one methylol group on each repeating unit. These methylol groups are available for crosslinking.

$$-N{=}C{=}O \quad + \; H_2O \; \rightarrow \; -NH_2 \; + \; CO_2\,(gas)$$

Figure 4

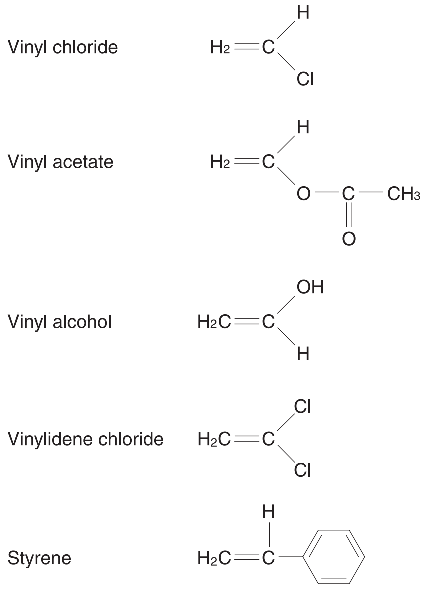


Figure 5

Amino resins This class of resins include melamine, urea and benzoguanamine (**Fig. 7**). The resins are mixed with formaldehyde to form a methylol group (see Phenolic resins) on the amino group. The methylol form of the resin is then exposed to a monohydric alcohol, which forms it into an alkoxy compound. During the baking (crosslinking) process, the alcohol group comes off of the amino resin and allows a covalent bond to be formed in a condensation process at the methylol site. These resins will crosslink at the following functional groups: $-OH$, $-COOH$ and $-CONH_2$.

Cellulose resins Cellulose resins are based on the natural cellulose polymer and, depending on which functional groups are present, the polymer could be nitrocellulose, cellulose acetate, cellulose acetate butyrate, ethyl cellulose or methyl cellulose. Cellulose resins are then combined with alkyds or oils to form lacquers.

Film formation

A coating is usually in liquid form when applied (except for powder coatings, see Application tech-

Phenol Formaldehyde

Figure 6

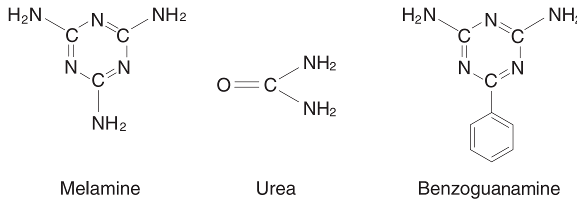


Figure 7

niques) but the final product desired is a solid. Film formation is achieved in four basic ways: evaporation of a solvent, polymerization, oxidation and coagulation.

Evaporation of a solvent As the name implies, all that is needed for this film former is time and circulation of air. Common types of evaporation finishes are acrylic lacquers and nitrocellulose lacquers. The final film of the solvent evaporation film formers are thermoplastic and convertible. Organic solvents are used, such as methyl ethyl ketone, toluene or mineral spirits, depending on the solubility of the primer and the intended application. Lacquer production is as simple as dissolving an acrylic resin in solvent. Film modifiers such as nitrocellulose resin or vinyl resin may be added to provide the final film with the desired qualities. Pigments may be added as a solid or in a resin solution.

Polymerization Polymerization is a polymer chain-forming reaction. It is used to form monomers into chains of polymers or in crosslinking two or more polymers together. This process is done by using heat or a catalyst. Two types of polymerization are addition and condensation (although oxidation, which may also be included in this section, will be addressed in the next section). Addition polymerization occurs when a free radical is formed and additional compounds are added with no by-products (**Fig. 8**). Condensation polymerization occurs when two compounds combine and, during this process, by-products such as water, hydrogen gas or carbon dioxide are evolved (**Fig. 9**).

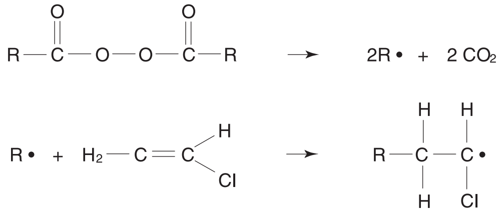


Figure 8

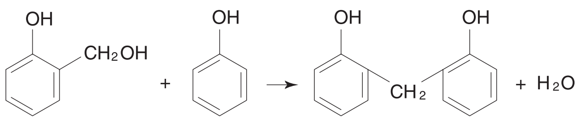


Figure 9

Oxidation Oxidation is the type of crosslinking that occurs in resins containing oils or fatty acids: varnishes, polyesters and alkyds. As stated above, oil-containing resins are classified as short, medium and long oil resins. The oils used are further classified as drying, semidrying, and nondrying oils, depending on the degree of unsaturation of the fatty acid chain. Drier additives may be added to aid the drying process. Dryers are heavy metal soaps of organic acids. The exact mechanism of the oxidation process is not known. What is known is that hydroperoxides are formed on the double bonds and that the dryers catalyze the decomposition of the hydroperoxide to the radical state. Some common oils used in oil-containing resins are linseed oil, tung oil, soya oil and coconut oil.

Coagulation This type of film formation occurs in systems that involve emulsion technology: latex (where water is the continuous or liquid phase) and nonaqueous dispersions (NAD) (where a solvent other than water is the continuous phase). In an emulsion, particles of the polymer are dispersed in a liquid in which they are not completely miscible. To keep the two semimiscible phases from separating, emulsifiers must be added. When the solvent evaporates, the polymers coalesce on a macromolecule scale. Common latex house paints are examples of this type of film formation.

End uses

Table 2, which is not exhaustive by any means, shows which resin systems may be found on different types of products.

Pigments, Fillers and Additives

Pigments, fillers and additives are in the liquid paint as solids. That is where the similarity ends. Pigments add color or metallic sheen, additives are used in small amounts to impart certain qualities on the paint or finished film, and fillers are generally used to add body or bulk to the paint.

Pigments

The four types of pigments to be discussed in this section are inorganic pigments, organic pigments,

Table 2 Some resin systems and their end uses

Resin system	*End use*
Epoxy amine	Marine Industrial
Epoxy esters	Automotive primers Maintenance paints
Epoxy phenol	Internal coatings for pipes
Epoxy amine formaldehyde (clear)	Brass Aluminum refrigerator trays
Epoxy acrylic enamel	Appliance Outdoor furniture
Short oil alkyds	Air dry primers and undercoats Primer on wood siding
Long oil alkyds	Exterior architectural paint Metal maintenance paints Topside marine paints
Medium oil alkyds	Automotive refinishing Traffic paints Implement enamels
Acrylic emulsions	Architectural paints (water-soluble) Automotive paint (NAD)
Acrylic emulsions in water with amino crosslinking polymers	Dip coatings Industrial baked finishes Waterborne basecoats (automotive)
Acrylic lacquers	Automotive refinishes Aluminum siding coating Aircraft finishes Wood finishes
Acrylic melamine enamel	Automotive
Phenolic varnishes	Floors Trim Can coatings Printing inks

filler and extender pigments, and metallic and pearlescent pigments. There are new specialty pigments coming on to the market for automotive topcoats, such as the color shifting pigments called Kameleon Kolors from DuPont. These specialty pigments will cause the observer to see different colors on the vehicle during 'travel'. Travel is a term used to describe the change in color when the painted surface is observed on its 'face' (parallel to the eyes) to the 'flop' (nearly perpendicular to the eyes). During the travel of the surface painted with the specialty pigments, the surface will actually change colors, depending upon the viewing angle. The specialty pigments are not commonly used yet, as they are expensive; therefore they will not be discussed in detail.

Inorganic pigments Inorganic pigments are the hiding or semihiding pigments, which simply means that these pigments have the ability to minimize substrate shade differences using the lowest possible film thickness. These types of pigment are also inexpensive but have low tinting strength, therefore more pigment is required to obtain a change in color. **Table 3** lists some inorganic pigments.

White and black pigments The most widely used white pigment is titanium dioxide (TiO_2). TiO_2 comes in two forms, rutile and anatase, which are crystal isomers. Rutile is the isomer of choice because it has 25% more tinting strength (the ability to tint to the desired color with the minimum amount of pigment) than anatase. Anatase TiO_2 is known to chalk or degrade when exposed to ultraviolet (UV) radiation. Rutile TiO_2 has a high refractive index (2.72) and that is what makes it such a good hiding pigment. The lead whites (basic carbonate white lead, basic sulfate white lead and basic silicate white lead) are not used, owing to the lead content. Two other whites are the zinc whites, zinc oxide and zinc sulfide, but the hiding power of the zinc whites is not as good as that of TiO_2.

There are four classes of black pigments: acetylene black, thermal black, furnace black and impingement or lamp black. Due to their particle size, good exterior durability and color, furnace and lamp blacks are used

Table 3 Inorganic pigments

Pigment	*Color*
White	See text
Black	See text
Chromates	
Lead	Yellow, green, orange
Zinc	Yellow, orange
Strontium	Yellow-green
Barium	Lemon yellow
Calcium	Yellow
Silicon	Orange
Oxides	
Iron	Red, yellow
Chromium	Light green
Lead	Red
Sulfides and selenides	
Cadmium	Orange, yellow
Mercury	
Ferrocyanides	
Iron	Blue
Copper	Maroon
Nickel titanate	Yellow
Ultramarine blue	Blue
Mineral violet (phosphate)	Purple
Krolor pigments	Orange, yellow, red
Cobalt-based pigments	Yellow
Fluorescent pigments	Red-orange

extensively in coatings. Other types of black pigment are available but are generally used for reasons other than color: graphite is used as an additive to impart chemical and mechanical properties to a film, vegetable and mineral blacks are used as fillers, etc.

Organic pigments Organic pigments are vibrant and very intense colors. They are costly and somewhat transparent but have a high tinting strength. Some organic pigments are listed in **Table 4**.

Many of the organic pigments can produce a number of colors, depending upon the functional groups or chelated molecules. An example would be phthalocyanine blue (**Fig. 10**). If the copper molecule (in the center) and 12–16 hydrogens are replaced with chlorine, the color of the pigment would be green.

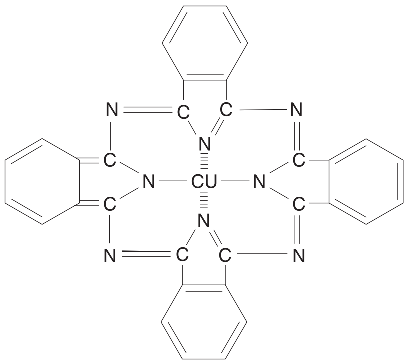


Figure 10 Phthalocyanine blue.

Filler and extender pigments Extender pigments were first introduced into coatings as fillers because they were lower in cost than the hiding pigments. However, as more knowledge was obtained regarding the effects that fillers have on certain properties of the finished film, 'fillers' became 'improvers'. Examples of fillers include, but are not limited to:

- *Calcium carbonates* Can be used to increase gloss, minimize fade in pastel colors and act as a fire retardant.
- *Calcium sulfate (gypsum)* Blended with TiO_2 to reduce cost and add whiteness. It is a soft pigment; therefore there is little wear on grinding equipment.
- *Barium sulfate (barytes)* This is an easy grinding pigment and therefore will be used in conjunction with harder-to-grind pigments to aid in their dispersion. Barytes has low water solubility and thus will have minimal effect on anions present in waterbased coatings.
- *Silica* Because of the hardness of this filler it will add abrasion resistance and has excellent non-chalking properties. Note: Silica comes in many forms: natural (crystalline, diatomaceous) and synthetic (pyrogenic silica, aerogels, precipitated).
- *Silicates (clays)* These will provide added 'scrubbility' and stain removal to latex paints, due to the flat, plate-like microstructure, and will improve the suspension of other pigments.
- *Magnesium silicates (talc)* In the fibrous form, talc replaces asbestos, which used to be used to resist mud-cracking and sagging. Talc will aid in overall structure of textured coatings.
- *Potassium-aluminum silicate (Muscovite mica)* The platelet structure helps to reinforce the film, minimize cracking, checking and vapor permeability, and is used as a base for the pearlescent pigments.
- *Calcium metasilicated (wollastonite)* This is used to stabilize emulsion systems and aids in hiding.

Table 4 Organic pigments

Azo pigments
Isoindolinones
Anthraquinones
Indigoid
Phthalocyanine
Basic phosphotungstic acid
Basic phosphomolybdic acid
Quinacridones
Dioxazines
Chelated nickel azo
Arylide red and maroon
Pyrazolone red
Perylene scarlets and vermillions

The aforementioned pigments and extenders are usually incorporated into a paint system by first producing a 'grind' of the pigment. A grind is made by mixing the dry pigment with a resin (and possibly a solvent) from the vehicle system that will be used in the final paint. The pigment is then pulverized in the resin using different types of mills (ball mill, sand mill, three-roll mill or Netzsch mill) or by using the shearing action of a special spinner blade. The pigments must be ground to a size that is smaller than the final film so that the pigment particles will not protrude out of the film. The exception to this is for pigments used in primers: protrusions are desired to aid in adhesion of the topcoat to the primer.

Metallic and pearlescent pigments Metallic and pearlescent pigments add sparkle and luster to a finish. The metallic pigments are generally aluminum flake dispersed in a resin or oil (the dispersion media

are proprietary), although gold bronze and zinc are also available. Pearlescent pigments are generally Muscovite mica flakes coated with TiO_2, iron oxide or other colored pigments. Pearlescent pigments may also be flakes of bismuth oxychloride, basic lead carbonate or natural pearl essence.

Metallic pigments There are two types of aluminum pigment: leafing and nonleafing. Leafing pigments float to the surface of the coating while it is still wet, producing a silvery metallic appearance. Nonleafing (the type used in automotive coatings) will disperse throughout the film, giving depth to the dry coating.

Gold bronze pigments are made by combining copper and zinc, atomizing them together and making flakes of the atomized particles in a ball mill. Zinc pigments are used in coatings at about 90% zinc by weight with an organo silicate or an epoxy ester. Extreme care must be taken to prevent the introduction of moisture into a container of zinc-containing paint because zinc plus water will produce zinc dihydroxide and hydrogen gas, which can cause an explosion in a closed container.

Pearlescent pigments The platelet structure of all the nacreous or pearlescent pigments gives them their rich appearance. Interference pigments is another name for pearlescent pigments. In the case of the TiO_2-coated mica flakes, the TiO_2 is deposited on to both sides of the mica, sandwiching the mineral between them (**Fig. 11**). The layers of TiO_2 are called 'interference layers'. Reflected light will bounce off the top and bottom interference layers and will, depending upon the thickness of the TiO_2 layers, reflect a certain color. On the other hand, if light is transmitted through the pigment, the color seen will be the complementary color to the reflected color. For example, if the reflected color is red, the transmitted color will be green. Iron oxide may also be used to coat the mica. These pigments will also act as interference pigments but will all have a red-yellow color, which is the absorptive color inherent to iron oxide. The resulting colors of the iron oxide-coated mica pigments are bronze to copper-red and will have a metallic luster. Combination pigments are pearlescence pigments that will have the coating of TiO_2 but will also have an additional layer of colored or absorptive pigment on the top and bottom of the TiO_2-mica pigment (**Fig. 12**). For instance, if iron oxide is the absorptive pigment and the interference color (from the TiO_2) is yellow, the color seen with reflective light is gold.

Bismuth oxychloride is a crystal that may be used alone to create pearlescent luster or can be deposited on mica flakes. Microscopically, basic lead carbonate appears as regular hexagonal crystals. These are being phased out owing to concern regarding lead. Natural pearlescence comes from crystals consisting of the purines guanine and hypoanthine, crystallized from fish scales and skins. Natural pearlescence is primarily used in cosmetics.

Additives

Additives are substances added to the paint to provide various properties. They include UV absorbers, hindered amine light stabilizers and antioxidants, which help protect the finish from the effects of UV radiation. Driers used in alkyd and other oil systems are additives. Wetting agents, such as carboxcylic acids and sulfuric esters used in the pigment-grinding process are also considered additives.

Solvents

The primary function of solvents is to aid in the application of the coating and to regulate the rate of evaporation, flow properties and viscosity. Solvents include water, hydrocarbon solvents, terpene solvents and oxygenated solvents. The forensic value of solvents is limited, as most paint received as evidence is in the form of a dry film. Properties that a paint chemist must consider when choosing a solvent are, among other things, solvency (the ability of a solvent to dissolve or disperse a resin), volatility (the evaporation rate), odor and toxicity. Due to more stringent environmental regulations, the use of volatile organic solvents is kept to a minimum. Reformulating the solvent portion of the paint, along with more efficient application techniques help paint manufacturers and production facilities comply with Environmental Protection Agency regulations.

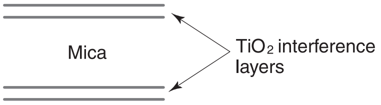


Figure 11 Pearlescent pigment.

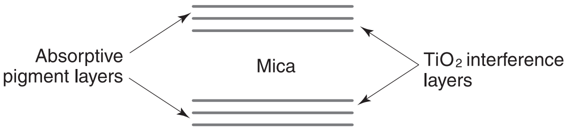


Figure 12 Combination pearlescent pigment.

Application Techniques

Coatings may be applied in many different ways. Brushing or rolling is a basic technique for applying a wet paint to a surface. There are various spray techniques, including air (an air stream propels the paint), airless (pressure propels the paint) and electrostatic (the paint is given a negative charge as pressure propels it on to a grounded substrate). All three of these methods atomize the paint and propel it on to a substrate. In the electrostatic spray methods the paint is actually attracted to the substrate and there is therefore less waste and easier clean-up (just wrap everything in aluminum foil!). For plastic parts, a conductive primer may be applied to the substrate.

Other forms of application are electrodeposition and powder coating. Electrodeposition 'electroplates' the substrate with paint where the substrate is either the anode or the cathode. The process occurs in the following manner: an object is dipped into a tank filled with paint and the current is activated. The paint then deposits on to the substrate, covering every conductive area of the surface with a uniform coat. The uniformity in film thickness is a property of the electrical potential used. The coated object is then usually placed in an oven.

Powder coating is an efficient method of depositing a coating on to a surface without the use of solvents. In this application technique the 'paint' is in the form of a powder manufactured by mixing a 'dry mix' of the resin, pigments, additives and fillers. These are then melted into a melt-mix. The melt-mix is then cooled, flaked and milled into a powder. The types of resin systems used can be thermoplastic systems, such as polypropylene, polyethylene, nylon and polyester, but polyvinyl chloride is most commonly used. The thermosetting resin systems used are epoxy, epoxy-polyester, polyester-urethane, polyester-glycidyl and acrylic-urethane. The application of the powder is achieved by using a fluidized bed system, electrostatic fluidized bed system and electrostatic spray. The fluidized bed system calls for two chambers stacked on top of one another: a plenum and a coating chamber which are separated by a porous plate (porous enough for air flow but not to let the powder escape). Compressed air is fed into the plenum and passed through the porous plate into the coating chamber to fluidize the powder. The substrate is then heated to a temperature slightly higher than the fusion point of the resin system used. When the substrate is exposed to the powder coating, the powder melts on to the substrate. In the electrostatic fluidized bed system the air in the plenum is ionized, it flows up into the coating chamber. In this system the substrate need not be preheated. Electrostatic spraying is the most widely used method. In common with conventional electrostatic liquid spraying, the particles are charged and the substrate is grounded, except that the coating is a powder instead of liquid. The powder that is not deposited on the substrate is reclaimed and reused.

See also: **Accident Investigation**: Motor Vehicle. **Analytical Techniques**: Separation Techniques. **Crime-scene Investigation and Examination**: Recovery of Human Remains. **Microchemistry**.

Further Reading

Allen RA (1981) *Epoxy Resins in Coatings*. Philadelphia: Federation of Societies for Coatings Technology.

Allyn G (1971) *Acrylic Resins*. Philadelphia: Federation of Societies for Paint Technology.

Brezinski D (1998) Lecture notes and information from the workshop 'Industrial Perspectives on Paint & Glass'.

Chenier P (1991) *Survey of Industrial Chemistry*. Malabar, FL: Krieger.

Del Donno T (1988) Waterborne coating outlook bright. *Industrial Finishing*. **December**:28–33.

Federation of Coatings Technology (1995) *SciQuest* vol. I. CD-ROM from Consolidated Research. Federation of Societies for Coatings Technology, Bluebell, PA.

Greenstein LM (1988) *Nacreous (Pearlescent) Pigments and Interference Pigments*, vol. I, 2nd edn. New York: Wiley.

Lewis P (ed.) (1988) *Pigment Handbook*, vols IIII. New York: Wiley.

McDonald R (1971) *Phenolic Resins*. Philadelphia: Federation of Societies for Coatings Technology.

Rice P and Werley W (1986) Modern powder coating manufacturing. *Industrial Finishing* **November**:23–25.

Rodriguez F (1989) *Principles of Polymer Systems*, 3rd edn. New York: Hemisphere.

Ryland S (1988) Lecture notes and information from the workshop 'Forensic Examination of Paint & Glass'.

Saferstein R (ed.) (1982) *Forensic Science Handbook*. Upper Saddle River, NJ: Prentice Hall.

Paternity Testing *see* **Deoxyribonucleic Acid**: Parentage Testing.

PATHOLOGY

Contents

Overview
Autopsy
Histopathology
Postmortem Changes
Postmortem Interval
Preservation of Evidence
Victim Recovery

Overview

D Ranson, Department of Forensic Medicine, Morosh University, Victorian Institute of Forensic Medicine, Melbourne, Australia

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0582

Introduction

In order to appreciate the role of pathology within the forensic sciences, it is important to understand how forensic pathologists are recruited and trained for the services that they provide.

Forensic pathologists are recruited from the medical profession, generally from among pathology specialists who practice in the field of anatomical pathology or histopathology. While the basic principles of pathology used in both clinical and forensic pathology specialties are the same, there is a considerable difference in the nature of the work performed by the different practitioners and in the mental and analytical processes they apply. The clinical pathologist's focus is on the application of diagnostic tests to enhance the clinical management of patients. In contrast, the forensic pathologist's focus is on the endpoint of the forensic investigation, which is usually a judicial process such as a criminal trial, coroner's inquest or other judicial or police inquiry. The legal system has long accepted that there is indeed an essential difference between those pathologists engaged in clinical, hospital-related practice and those engaged in forensic pathology practice.

In contrast to the clinical pathologist, the forensic pathologist has to be trained in, and develop an understanding and aptitude for, the legal process of investigation. In addition he or she must be able to integrate this legal approach with the scientific and medical process of investigation. These two approaches to investigating issues are distinctly different. Traditionally, medical practitioners have had little or no understanding of the legal investigative method as it is not part of the basic medical curriculum.

A practical result of this difference is that forensic pathology does not sit comfortably within a professional clinical pathology environment. Indeed, the service requirements of forensic pathology mean that the service is usually more efficiently provided by specialist forensic pathologists working together and providing a broad-based regional forensic pathology service for a range of clients. Such an organizational structure allows for:

- continuity of professional services;
- internal case review and audit;
- personal professional development;
- concentration of forensic expertise;
- support for undergraduate and postgraduate training;
- integrated research and development opportunities.

In some jurisdictions, this full time forensic medical service is the preferred organizational structure for forensic pathology, while in others this structure is combined with part-time practitioners who may also have a clinical pathology role.

In many jurisdictions outside the English legal system the distinction between clinical forensic medicine and forensic pathology is not made and practitioners perform both forensic roles, often operating from specialist forensic medical institutes that have their origins in the public health system.

The nature of forensic pathology is such that the service does not require a large number of medical practitioners in any particular jurisdiction. However,

a continuous supply of a small number of experienced and well-trained medical practitioners is needed to provide for ongoing forensic pathology services. In this regard, undergraduate and postgraduate teaching, together with affiliation with medical schools and the appropriate professional training colleges, is an important part of any forensic pathology service.

Legal jurisdictions differ in the way they organize death investigation services and the relationship between a coroner and police investigators is at times complex. In Australia and England and Wales the authority to investigate a death lies with a coroner and it is the coroner who authorizes a pathologist to perform a postmortem examination. The police have a similar duty to investigate a death and, with respect to certain aspects of the death investigation, they act as agents of investigation for the coroner. In this situation, the forensic pathologist is another coinvestigator for the coroner.

Where a matter involves direct criminal issues, much of the practical investigation is carried out by the police on behalf of the criminal justice system, a system in which the coroner plays little part. In these circumstances, there is a practical basis for concluding that the pathologist is in fact carrying out a service for the police investigators. In reality, a criminal investigation involves many phases. Traditionally, the forensic pathologist was involved in that phase of an investigation that is centered on a death or injury, and in particular the examination of an injured or deceased person to ascertain the nature and cause of the injuries. In practice, however, forensic pathologists play a wider role with respect to the criminal investigation. For example, they may become involved in the examination of scenes of death, or in the examination of suspects who may have inflicted injuries on the victim. They may evaluate medical records for medicolegal purposes, and examine the statements of other witnesses with regard to medical matters.

Forensic pathologists have a wider role than just dealing with suspicious deaths. The vast majority of autopsies conducted by forensic pathologists involve unconfirmed natural deaths which society requires to be scrutinized and confirmed. The investigation of nonsuspicious natural deaths, accidents (fatal incidents), deaths from suicide, and deaths from other forms of intentional and nonintentional injury form the vast majority of their case work. The information that can be gained from investigating these deaths is potentially very significant for the community. The role of the forensic pathologist in relation to suspicious deaths and the criminal justice system is the one most readily understood and appreciated by the community. However, the wider role with respect to autopsies in nonsuspicious cases is far less defined and less understood by the lay, legal or even the general medical community.

Training

The training of a forensic pathologist takes into account the fact that forensic pathology is based on the techniques and skills of clinical pathology, in particular anatomical pathology, histopathology and morbid anatomy. Any substantial period of study and training will equip a clinical pathologist with the additional factual knowledge of forensic medicine and science that is necessary in order to deal with the scientific investigatory issues that surround the work. The development of the necessary legal investigatory skills is more difficult. In the past, the majority of forensic pathologists gained these skills by experience. The very act of taking part in complex police investigations and in court processes within the coronial and criminal legal systems gradually provides forensic pathologists with an insight into the legal investigatory process and the evidential requirements of the legal system.

There have been several attempts to introduce training in the legal method into some undergraduate and postgraduate forensic medical courses. For example, in a number of undergraduate courses in forensic medicine students take part in moot (practice) courts and a number of joint law–medicine subjects are now available for undergraduates from faculties of law and medicine. Despite this, it is not easy for medical students to acquire knowledge and understanding of the legal method. Increasingly, medical practitioners who are interested in forensic medicine and forensic pathology are turning to the study of law itself to acquire these skills.

With regard to postgraduate training, there has been a considerable increase in the interest of the professional pathology organizations around the world in the specialty area of forensic pathology. The Royal Colleges of Pathologists of both Australasia and the United Kingdom offer specialist fellowship and membership examinations in forensic pathology. While accrediting examinations are now available in forensic pathology, training programs are still limited. The degree of organization varies from place to place, and the lack of formal training programs has been recognized in a number of jurisdictions. The importance of formal training and accredited external examinations cannot be overemphasized. If society and the courts are to be assured that they are receiving expert opinion evidence from a person who has the necessary skills and knowledge, then the availability of accredited qualifications based on organized training is essential.

The Royal College of Pathologists of Australasia in its information booklet for trainees states that forensic pathology trainees must gain the following experience/knowledge:

1. Conduct under supervision of at least 500 coronial postmortem examinations comprising a mix of adult and infant deaths and unexpected natural deaths, accidental deaths, suicides, homicides and postoperative and anesthetic deaths.
2. Training and experience in special autopsy procedures relevant to forensic pathology, including:
   (a) the demonstration of the vertebral arteries, the cervical cord, the urethra and structures of the neck;
   (b) enucleation of the orbits;
   (c) neonatal and infant autopsy techniques;
   (d) identification techniques;
   (e) the collection of organs, tissues and fluids for toxicological examination.
3. A working knowledge of associated forensic fields including:
   (a) toxicology and its place in forensic pathology;
   (b) forensic radiology;
   (c) forensic odontology;
   (d) osteology;
   (e) forensic immunohematology.
4. Attendance at scenes of suspicious and homicidal deaths, including training in the principles of trace evidence collection and its preservation at the scene of death and at the postmortem examination.
5. The use of biological and physical forensic sciences in assessing the mechanism and cause of accidental, homicidal and suicidal deaths and in reconstruction of the circumstances surrounding such deaths.
6. Giving factual and expert evidence in Australian courts of law.
7. The functions, operation and legislation relevant to the coronial and criminal justice systems, and the law relating to transplantation and the determination and certification of death in Australia and New Zealand.

To qualify as fellows of the Royal College of Pathologists of Australasia in forensic pathology, candidates sit a final written examination which covers both anatomical pathology (morbid anatomy) and forensic pathology. The final practical examinations include assessment of microscopic sections of human tissues in selected pathology cases, both anatomical and forensic, and forensic case analysis based on the examination of photographs of crime and death scenes and autopsy findings. The fellowship final examination also requires the candidate to present a case book in which 10 coronial postmortem examinations are described and annotated in detailed discussion papers. It also includes a viva voce in which candidates assess a variety of visual material as well as answering formal questions. Before the final examinations can be attempted, candidates must have already passed preliminary examinations in basic pathological science and anatomical pathology, which include both written and practical tasks.

The aim of professional examinations is to insure that the candidate has reached a standard of proficiency which the profession recognizes as being the minimum required to undertake competent professional practice in that field. The acquisition of such professional qualifications does not exempt the pathologist from further study. Continuous professional education is now a requirement of most specialist branches of medicine; many of the Royal Colleges governing professional practice require fellows and members to engage actively in educational activities in order to retain their professional accreditation. While courts require expert witnesses to prove their expertise each time they give evidence, a witness who holds professional qualifications that carry with them the obligation of continuing professional education is more easily able to prove that his or her knowledge and expertise is up to date.

Forensic pathology is not a discipline that stands alone as an unconnected and discrete specialty. In practical terms, forensic pathology is a subspecialty of pathology; most practising forensic pathologists have a wide range of experience in the areas of clinical pathology, and in particular the subspecialties of anatomical or histopathology.

Specialist Skills

As a result of their training, forensic pathologists acquire a variety of specialist skills that they apply in their work. The basis of these skills is completion of a medical course, including appropriate clinical internship and a variety of other clinical appointments. As a result, the skills of a forensic pathologist cover a wide variety of subdisciplines in medicine, and particularly subspecialities within pathology. These skills can be divided into clinical, pathological and legal areas, together with general skills in science and communication.

Clinical skills

Forensic pathologists are medical practitioners who have undergone a full medical course, of 5–7 years, which has included both scientific preclinical and clinical studies. They will usually have spent several years working in clinical and or diagnostic medicine within a hospital or community setting. However, it should not be assumed that because of such experience they are currently experts in clinical medicine.

Forensic pathology skills

As mentioned above, forensic pathology involves the application of basic pathology disciplines in the forensic or medicolegal setting. The pathological skills involved in forensic pathology can be divided into a number of areas: anatomical pathology, neuropathology, cytopathology, hematology, microbiology, immunology, and chemical pathology and toxicology.

Anatomical pathology Pathologists practicing in these areas have skills in the macroscopic or naked-eye examination of diseased organs and tissues, and microscopic examination of human organs and tissues. Microscopic pathology examination, sometimes referred to as surgical histopathology, forms the bulk of the work of anatomical pathologists in a hospital setting. The anatomical pathologist also performs autopsies in a hospital setting where consent has been obtained from families and next of kin.

Most forensic pathologists have completed full anatomical or surgical pathology training and therefore they are experienced in the performance of hospital autopsies as well as forensic pathology autopsies. In addition, they are skilled in the areas of microscopy of human tissues and the identification within those tissues of various types of human disease.

Neuropathology Is a discrete subspecialty within anatomical pathology, neuropathologists deal with diseases of the brain, the spinal cord, the peripheral nerves and the muscles of the body. Neuropathologists also deal with the interpretation of traumatic damage to the tissues of the central nervous system. The skills of neuropathology have a particular relevance for forensic pathology with regard to the processes involved in head injuries.

Cytopathology Cytopathologists examine the cells of the body in isolation, or in small clumps where the cells do not form part of an intact tissue structure. The cells are obtained through a variety of sampling processes: scraping of the surface of tissues, as in the examination of the cervix or neck of the uterus; aspirating fluids from various parts of the body; or aspirating solid tissue masses. The role of cytopathology within forensic pathology is limited. However, a wide variety of cytological techniques are employed within forensic pathology and some tests for drowning and for the identification of spermatozoa employ procedures similar to those used in cytopathology.

Hematology Hematology involves the diagnosis and treatment of diseases of the blood and includes the examination of peripheral blood and the blood precursor cells found in the bone marrow. While the clinical aspects of hematology do not impinge greatly on the work of the forensic pathologist, many of the principles of the identification of blood and the determination of blood grouping used in forensic science and forensic medicine involve the same laboratory techniques and procedures.

Microbiology This branch of pathology deals with the identification of microorganisms that cause disease. Bacteria, viruses and fungi are just some of the agents that are dealt with by a microbiologist. Like other pathologists, the microbiologist deals with samples taken from the human body but, in addition, their work can involve the analysis of specimens taken from the environment. These samples are examined to determine which microorganisms are present, and often the organisms are tested to see whether they are sensitive or resistant to antibiotics and other drugs.

Microbiological diseases play a part in a number of deaths investigated by forensic pathologists, and many of the individuals whose deaths form the subject of forensic pathological investigation have lifestyles that involve current or previous infection with particular agents. A good example of this is death associated with overdose of intravenous drugs, where there is often prior infection with *Hepatitis B virus, Hepatitis C virus* or *Human immunodeficiency virus* (HIV).

Immunology Immunology has grown as a division of pathology and is now recognized as a major subspecialty. Immunologists study and test the function of the immune system of the body and the diseases that are associated specifically with immune system dysfunction. From the perspective of forensic pathology, many of the techniques of immunology are used in the forensic testing processes. The serological tests and blood grouping tests that are a feature of forensic science and forensic medicine are based on immunological principles.

Chemical pathology and toxicology Sometimes referred to as medical biochemistry, chemical pathology can also encompass the field of toxicology. The field of toxicology uses many of the techniques of

analysis that are found in the chemical pathology laboratory. The toxicologist does not generally measure natural body substances but analyzes human tissues for the presence of drugs and other chemical agents that may have been taken into the body. Many chemical pathology departments analyze body samples for the presence of drugs, and in fact provide a limited toxicology service, usually to support medical treatment and diagnosis of poisoning or adverse drug effects. From the perspective of forensic pathology, chemical pathology and toxicology is an important related discipline. Toxicological analysis is a routine part of most forensic autopsies. Homicides, suicides and motor vehicle accidents are perhaps the most common cases in which drug analysis is involved. However, there is a wide variety of apparently accidental deaths, including those associated with work and recreation, where toxicology and drug analysis are important in reconstructing the circumstances of the death. As a result of this, forensic pathologists are regularly required to incorporate the results of toxicological analysis in the conclusions that they reach regarding the autopsy findings and cause of death.

Legal skills

The area of knowledge that distinguishes forensic pathologists from their clinical colleagues is their understanding of legal process, medical law, court procedures and rules of evidence. While doctors who engage in civil injuries work are also familiar with court processes, the remainder of the medical profession has little contact with the legal system and consequently has little knowledge of the legal principles involved in civil and criminal cases. Many forensic pathologists have gained their knowledge of the legal system through long experience of working with it.

The training of medical practitioners is based on the scientific method and the scientific investigative process. While the legal system uses the scientific investigative process in some areas, its operation relies on a legal investigative process that differs both in structure and philosophy from the theories of scientific investigation. It is the appreciation of this that distinguishes forensic medical practitioners from their non-forensic colleagues. An understanding of the nature of the legal investigation process and the role of the medical practitioner within it, both before and during a judicial hearing, enables the forensic medical practitioner to play an effective part in the judicial process.

Scientific skills

When assessing the skills of the forensic pathologist, it is important to remember that, as doctors with a general background and training in medicine, their early training both at secondary and tertiary level has been based on the sciences. All medical students study science, either before or at the time of entering a medical course, and in many medical schools the first year of undergraduate medical teaching often includes further study of physics, chemistry, mathematics and biological science. The remainder of undergraduate preclinical and clinical training relies upon the knowledge of scientific principles, and such basic scientific knowledge is further developed and reinforced during a medical course.

Communications skills

Medicolegal communications includes the delivery of oral evidence as well as the compilation of intelligible medicolegal reports. Skill in this area is a particular attribute of forensic pathologists. Through their undergraduate training and additional years of postgraduate training, doctors develop and rely upon communication skills, often referred to colloquially as ‘bedside manner’. Doctors are familiar with techniques of overcoming barriers to communication. They are aware of the deliberate use of silence, and of confrontational and summarizing techniques. When one examines these clinical communication skills, it becomes clear that they are similar to the skills of barristers examining or cross-examining witnesses in court. In practical terms, the members of a jury are very similar to patients in a medical consultation. Their range of medical knowledge, intellectual ability, scientific background, attention and commitment mirrors closely the range found by doctors when dealing with patients. As a result, the skills that a doctor uses to communicate with a patient can be used to great effect in communicating with a jury.

Roles

We have looked at the skills acquired by forensic pathologists in their training and experience. In what ways then do forensic pathologists utilize their skills in everyday practice? In order to understand this, we need to identify the various roles that forensic pathologists take in their work and the services they provide. Forensic pathologists, like many medical specialists, in fact practice a number of roles: medical practitioner, public health practitioner, death scene examiner, dead body examiner, forensic biologist, forensic toxicologist and medical detective.

Medical practitioner

Most forensic pathologists do not take part in clinical medical practice involving the diagnosis of illness and disease or the provision of surgical and medical

treatment; however, as medical practitioners they are still committed to the maintenance of good health within the community. The forensic pathologist retains the basic medical skills in clinical diagnosis and therapeutics. In providing an autopsy service for the coroner and other individuals involved in the health and justice systems, the forensic pathologist is involved in analysis of clinical, diagnostic and therapeutic issues as well as broad public health issues, including occupational health and safety. The media, the lay community and indeed a number of medical and legal professionals tend to view the forensic pathologist as being concerned only with criminal justice aspects of death investigation. In reality, deaths involving a suspicion of criminal activity represent a small percentage of the work of the forensic pathologist. While the forensic pathologist remains a watchdog on the lookout for criminality during death investigations, most of their work involves deaths associated with natural disease and unintentional traumatic injury. These deaths have important considerations for public health policy and health service planning as well as for community health and safety.

Public health practitioner

Community health and public health are subspecialties in the area of clinical medical practice. Forensic medicine, particularly in Europe, had its origin in the work of the public physicians, doctors who had responsibility for public health and general public medical services in city states during and after the Renaissance. The justice system turned to these early public health physicians in order to obtain medical services to assist in the investigation of crimes, particularly crimes against the person. In time, the office of public physician formed the basis of the public and university institutes of forensic medicine that developed in Europe. These European developments were also seen in Scotland, where the early university departments of forensic medicine were joint departments of public health. This early association of forensic medicine and pathology with public health was gradually lost, but forensic medical institutes are returning to this field by becoming involved in issues of public health, occupational health and community health.

Death scene examiner

The case investigation for a forensic pathologist is initiated by a death; the investigation process includes analysis of information relating to the deceased person prior to their death, together with examination and analysis of the scene and environment in which they died or were found dead.

In allegedly suspicious deaths, the death scene is examined by a team of individuals that comprises police officers, forensic scientists, forensic pathologists and other specialist examiners, such as forensic anthropologists (for investigation of skeletal remains). A variety of other individuals are also involved in the examination of a scene of a suspicious death, including photographers and video camera operators.

In most cases after a dead body is found in suspicious circumstances, the forensic pathologist is called to the scene, not only to examine the body and provide initial information to the investigators but also to study the environment in which the body lies. The investigation of the environment of the death helps the pathologist in coming to conclusions regarding the subsequent macroscopic and microscopic autopsy findings. In many cases, by virtue of training and long experience in attending at death scenes associated with crimes, the forensic pathologist can contribute directly to the crime scene examination and provide initial advice in areas of forensic science, including forensic biology and ballistics. The inclusion of specialist forensic scientists from a wide variety of disciplines in the death scene investigation team is important, and the forensic pathologist may well have a part to play in helping to determine the appropriate specialists that may be required.

Dead body examiner

The media and most of the lay public recognize the forensic pathologist as a dead body examiner. In many cases this involves the pathologist performing a preliminary examination of a body at the death scene (or the place in which the body is found). The more detailed formal examination of a dead body occurs during the process of the autopsy. But the autopsy is more than an external examination of the body followed by dissection and an internal examination. The process of an autopsy covers a wide range of activities, and can include radiology and fluoroscopy as well as endoscopic techniques for examining the interior of the body without formal dissection.

The forensic autopsy is the area of death investigation for which the forensic pathologist has sole responsibility. The provision of timely autopsy information regarding causation of injuries and cause and time of death to co-investigators, such as the police, may result in better focusing of police investigation.

Forensic biologist

Forensic biology is one of the core divisions of forensic science. Forensic biology services are for the most

part provided through major institutes of forensic science, and a number of these exist in association with institutes of forensic medicine and pathology. Given the concentration upon basic biological principles in undergraduate medical training, it is not surprising that a forensic pathologist often has the necessary expertise to take part in forensic biological service work and to give expert evidence in the field of forensic biology.

Today, the forensic application of developments in molecular biology involving the analysis and comparison of DNA has been one of the great success stories of forensic biology. Forensic pathologists are often well equipped to analyze and comment on the general use of these techniques.

Forensic toxicologist

Medical pharmacology and therapeutics form part of the basic education of all legally qualified medical practitioners. Knowledge of drugs and their normal and abnormal effects on the human body are fundamental to medical practice. While most doctors understand the toxic effects of therapeutic and some nontherapeutic drugs, details regarding the wide range of substances that are dangerous to the human body exceed the knowledge of most doctors. It is scientists specializing in forensic toxicology who screen and analyze human tissues for drugs and toxins. Forensic pathologists have a specific interest in this area. This is because the forensic pathologist has to determine issues relating to the cause of death and circumstances of death of individuals. Where such deaths are apparently the result of toxic substances, then a forensic pathologist must collect the appropriate body samples and, together with a forensic toxicologist, arrange for the most appropriate analysis.

Medical detective

Forensic pathologists, more than all other specialists in medical practice, are medical detectives. They are involved in all phases of a death investigation and form part of the detection team that analyses the circumstances of a death. This role of medical detective is often glorified in the media, where the forensic pathologist is frequently elevated to the key player in the investigation. In reality the forensic pathologist participates on an equal basis with all other specialist investigators in the team investigating a death.

Forensic pathologists provide far more than an autopsy service for the justice system. Their unique range of skills and expertise overlaps many of the traditional scientific, medical and legal compartments, and as such can be a highly effective and efficient resource for investigators.

See also: **Education**: An International Perspective. **Toxicology**: Overview. **Legal Aspects of Forensic Science**:

Further Reading

Brodrick Committee Report (1971) Cmnd. 4810 *Report of the Committee on Death Certification and Coroners*. London: Home Office.

Gonzalez-Crussi F (1987) *Three Forms of Sudden Death: and Other Reflections on the Grandeur and Misery of the Body*. London: Picador.

Hill RB and Anderson RE (1988) *The Autopsy: Medical Practice and Public Policy*. Boston: Butterworth.

Knight B (1991) *Forensic Pathology*. London: Edward Arnold.

Plueckhahn VD and Cordner SM (1991) *Ethics, Legal Medicine and Forensic Pathology*, 2nd edn. Melbourne: Melbourne University Press.

Selby H *The Aftermath of Death: Coronials*. Leichhardt, NSW: Federation Press.

Spitz WU and Fisher RS (1993) *Spitz and Fisher's Medicolegal Investigation of Death: Guidelines for the Application of Pathology to Crime Investigation*. Springfield, IL: Thomas.

Autopsy

D J Pounder, University of Dundee, Dundee, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0584

Introduction

The autopsy, or necropsy as it is sometimes known in Britain, is a postmortem dissection of the body. The literal meaning of autopsy is 'to see for oneself'. Autopsies are of two types: those performed in a hospital setting for purely medical purposes, and those performed under legal authority for forensic purposes. The 'hospital' autopsy and the forensic autopsy address different questions and have different approaches. These differences are reflected in the conduct of the examination, the ancillary investigations, the significance attached to physical evidence, the approach to documentation, and the requirement for chains of custody. The 'hospital' autopsy focuses on the internal examination of the body and the correlation of the findings with the clinical record. The broad purpose of the medicolegal autopsy is to gather evidence relevant to the investigation of a

death by legal authorities. The medicolegal autopsy has the purpose of establishing a cause of death; assisting in interpreting and correlating the facts and circumstances surrounding the death; recovering and preserving evidence; reconstructing how injuries occurred; establishing the identity of the deceased; estimating the time since death; providing an accurately observed record of the medical findings; and assisting in addressing the medicolegal aspects of public safety, public health and criminal and civil court proceedings.

Autopsy Strategy

A medicolegal death investigation is much broader than the autopsy, and includes the scene of death examination and the gathering of witness and documentary evidence. The latter includes the social, psychiatric and medical history of the deceased. This background information informs and directs the conduct of the forensic autopsy. Every autopsy begins with a review of the available investigative information, an evaluation of the medicolegal issues arising, and the development of a strategy for the conduct of the autopsy. This strategy draws upon a repertoire of techniques, and structures the conduct of the postmortem examination to obtain the maximum amount of evidence relevant to the specific issues of the case. Such a strategy is inevitably flexible, as the ongoing examination may bring new information to light.

Identification

Who, when, where, why, how and what are the questions that the autopsy assists in answering. Who has died is not usually at issue, but establishing identity can be a problem in some instances. There may be no circumstantial evidence of the identity of a person found dead, or visual identification may be impossible because of decomposition, fire damage, physical disruption or mutilation. In all of these circumstances special techniques will be required and then identification becomes a multidisciplinary effort. Radiology, dentistry, anthropology, law enforcement, criminalistics and molecular biology may be involved. The commonest methods of scientific individualization are fingerprints, dentistry and radiography. DNA analysis is increasingly used. The essential principle of individualized identification is to make a match between a unique feature, physically part of the body, and an antemortem record of that unique feature. This requires an accurate and reliable antemortem record. In exceptional circumstances it may be necessary to resort to the less robust methods of photographic superimposition and facial reconstruction.

Time and place of death

When the person died is usually best established from the circumstantial evidence, and is not commonly an important issue. However, it may be critical when alibi is the defence in a homicide, or when the sequence of deaths of family members killed in the same incident influences the inheritance of estates, or when life insurance policies have lapsed in the time between the disappearance of a person and the discovery of the body. Again, special techniques need to be applied to establish the time since death. Where the person died is usually where the body was found, but not always. Homicide victims may be dumped, and bodies may travel long distances in rivers and the sea prior to their recovery. The autopsy may provide evidence that the body has been moved after death, as well as trace evidence originating from the place of death.

Cause of death

Establishing why the person died – that is, the disease or injury initiating the sequence of events, short or prolonged, leading to death – is one of the most important functions of the autopsy. However, the cause of death may be readily apparent from the circumstances and the condition of the body, such as a passenger in an air crash, and in these cases the autopsy is required for other investigative reasons, such as victim identification and accident reconstruction. There are many causes of death, both natural and unnatural, that cannot be established with certainty by autopsy alone. Examples include epilepsy, asthma and drowning. In such deaths the autopsy assists by excluding other potential causes of death from trauma, drugs or natural disease. In a forensic autopsy, the demonstration of the absence of injury or disease may be as important as the demonstration of unequivocal or occult injury. Some natural diseases, particularly coronary artery disease, are chronic and can kill at any time, so their identification at autopsy represents a potential cause of death but not necessarily *the* cause of death. Similarly, drugs such as alcohol, morphine and other opiates, to which individuals develop tolerance, may be found at autopsy in concentrations in blood sufficient to account for death but not necessarily lethal. Very minor injuries may be consistent with a subtle form of homicidal trauma, such as soft smothering, and sufficient to arouse suspicion but not provide conclusive proof. Only in a minority of deaths are there autopsy findings of a trauma, poisoning or natural disease incompatible with life, so that the cause of death is established with certainty by the autopsy alone. For the great majority of death investigations, establishing why the person died requires the integrated analysis of the autopsy findings and

scene of death and anamnestic data. In a small number of deaths a thorough investigation, including an autopsy, may fail to establish the cause of death. Such a case is characterized as a 'negative autopsy' in the English-speaking world and as a 'white autopsy' in the Spanish-speaking world.

Manner of death

How the person died – that is, the manner of death – is a single word description of the circumstances leading up to the death. The manner of death may be natural or unnatural, and if unnatural may be accidental, suicidal or homicidal. Deaths from alcohol and drug abuse are difficult to classify and in some jurisdictions are simply described as 'unclassified'. For some deaths the manner may be undetermined, either because the cause of death is unknown or because the circumstances are unclear; for example, whether a drowning was accidental or suicidal. Responsibility for determining the manner of death always lies with the legal authority, but this may be the autopsy pathologist if the jurisdiction has a North American-style medical examiner system.

Surrounding circumstances

Establishing what happened to the deceased – that is, the totality of the events leading up to and surrounding the death – is invariably the focus of any legal proceedings. The autopsy can assist in this by establishing objective evidence of some of the things that happened or, more commonly, evidence limiting what could have happened. In homicide investigations, the autopsy can provide information on victim position, weapon used, sequence of injuries, postinjury activities, period of consciousness, survival time, time of death, postmortem interference with the body and method of disposal.

Legal Authority

The law, regulations and local administrative practices governing medicolegal autopsies vary from place to place. Wherever the jurisdiction, before the start of any autopsy there must be legal authority to proceed. Only the appropriate law officer can provide this authority, and he or she may be a judge, prosecutor, coroner or medical examiner. The consent of the next of kin is not required, and in most instances they have no effective means of objecting. However, ethnic and religious minorities who have principled objections to postmortem dissections have achieved the legal right of objection in some jurisdictions. The pathologist receiving the authority for autopsy should permanently record how it was received, from whom and when. The autopsy report will ultimately need an introductory statement of the legal authority upon which it was performed. The next requirement is a formal identification of the decedent by a method meeting the necessary legal requirements. In a case in which no court proceedings are anticipated, simply noting the details contained on the mortuary body tag and any accompanying documentation may suffice. In other circumstances visual identification of the decedent by next of kin, possibly in the presence of a pathologist, or the use of a scientific method may be required. Whatever method is used, it should be formally recorded in the autopsy report.

External Examination

The medicolegal autopsy requires the accurate observation and accurate documentation of both the external and the internal examination findings. The external examination is an essential component of the postmortem examination and constitutes the first part of the forensic autopsy. In some deaths it may be appropriate to limit the postmortem examination to the external examination without internal dissection. This is a judgment to be made by the pathologist and law officer in light of the legal requirements, public health concerns, potential for criminal prosecution or civil litigation, resource constraints, and the religious and personal views of the next of kin. The external examination embraces everything on and upon the body, including clothing, physical evidence and medical paraphernalia.

Clothing

Examination of the undisturbed clothing is a valuable part of the autopsy. In some jurisdictions the body is stripped naked by police officers or mortuary technicians before presentation to the pathologist, but this is an unwise practice. The clothing can provide a wealth of useful information on the lifestyle of the decedent, events surrounding the death and the cause of death. If the pathologist does not document the clothing then it is often not documented at all. Each article of clothing should be described in appropriate detail and, when the body is unidentified, details of the labels and laundry marks as well. Description of the clothing should include general descriptions of any disarrangement, damage and stains. Recovery of trace evidence from clothing may be undertaken either at the scene of death or in the autopsy room, depending upon local practice and the nature of the case. Trace evidence might include hairs, fibers, paint chips, glass fragments, vegetation and insects. The collection and storage of this trace evidence must meet the legal

requirements for the chain of custody. The appropriateness of the clothing should be assessed against the scene of death and anamnestic information, particularly in potential hypothermic deaths. Stains, scuffmarks and tears to clothing may assist in traffic accident reconstruction or in clarifying events surrounding a death. Gunshot holes and stab wounds to clothing provide useful information in themselves, but more so when correlated with the underlying injuries to the body. Bloodstain patterns to clothing may illuminate the events following trauma and the activities of the victim prior to collapse. Jewelry may provide evidence of identification; medical bracelets and necklaces may indicate a chronic disease; pockets may contain medication or drugs of abuse; and personal papers may give information on identity, medical history and lifestyle.

Stains

Before removing clothing and personal effects, and cleaning the body of any stains, it may be necessary to make a permanent photographic record, particularly in a case of homicide. After removal of the clothing, a head-to-toe detailed examination of the naked body is made. In the first instance this external examination should document stains and soiling, general and specific identifying characteristics, evidence of medical intervention and postmortem changes. Stains to the exposed body surfaces may be described at the time of the examination of the clothing and supplemented with a description of stains to the whole body. The location, extent and type of staining or soiling can provide useful information. Blood flow patterns from wounds reflect body position after wounding; high-velocity impact blood splatter and gunshot residue stains on the hands may indicate suicide; 'coffee grounds vomitus' around the mouth and melena staining of the buttocks suggest death from massive gastrointestinal hemorrhage; vomitus containing tablet debris raises the possibility of suicidal overdose.

General descriptors

The general external description should include height, weight, build, sex, race, head and body hair, eyes, dentition, scars, tattoos and body piercings and evidence of natural disease, particularly ankle edema and varicose veins, which raise the possibility of a cardiovascular death. The back, anus and perineum, palms and soles must always be examined. A specific search for petechial hemorrhages in the eyelids, conjunctivae, inner lips, face and neck is mandatory because they are easily overlooked but of considerable importance in the diagnosis of asphyxia.

The recording of postmortem changes to the body does not usually include body temperature unless there is a specific concern about the time of death. Core body temperature obtained rectally, or by a liver stab, is required. The presence or absence, and the pattern, of postmortem lividity, rigor mortis and putrefaction are routinely observed and recorded. Postmortem lividity (livor mortis or hypostasis) reflects gravitational pooling of blood after death, and thus body position. Areas of contact pallor produced by pressure from clothing or adjacent objects should correlate with the scene of death findings. A pink lividity, rather than the usual purple-red, raises the possibility of death from carbon monoxide poisoning, cyanide or hypothermia. Rigor mortis, which develops some hours after death, fixes the body in the position in which it came to rest and should also correlate with the scene of death findings. Neither lividity nor rigor is of any substantial value in estimating time of death. Postmortem injuries produced during the recovery of the body or by the feeding of insects, birds, animals or crustaceans are recorded separately from injuries produced in life in order to avoid confusion.

Medical intervention

In many deaths there is evidence of attempts at lifesaving medical intervention. Where disposable medical equipment is attached to the body this should not be removed at the scene of death but transported with the body to the mortuary to be recorded by the pathologist. The most common items are airways and solutions for intravenous infusion. All of these require description and an assessment of their correct placement, together with recording of any associated bleeding, bruising or other tissue damage. Emergency medical treatment is rarely documented in detail at the time it is given because of the urgency of the circumstances, and consequently the autopsy record is often the most complete and reliable record. It may be important in any civil litigation for malpractice. Injuries produced by medical intervention, particularly those in the neck, may be misinterpreted as assaultive if not viewed in context. Recording intravenous lines eliminates the possibility that the associated needle puncture marks may give rise to a false suspicion of intravenous drug abuse.

Injuries

The final stage of the external examination is the documentation of injuries. These are described systematically, either by grouping them according to injury type and anatomical location, or by numbering them, without implying an order of infliction or ranking of severity. Each injury is characterized by its type

– for example, bruise, abrasion, laceration, incised wound, stab wound, gunshot wound, burn – and its general anatomical location. The precise anatomical site of an injury is recorded in cases of homicide or if it is of particular significance for the reconstruction of the circumstances – for example, a single suicidal gunshot wound or an imprint injury from a vehicle striking a pedestrian. Precisely locating a wound is analogous to giving a latitude and longitude with respect to fixed landmarks, which may be the midline of the chest, the heel or top of the head, or any fixed bony prominence. The size, shape and other relevant features of the injury, depending on its type, are observed and recorded. For this purpose photography is of special value in documenting a wealth of detail. The individual external injuries are described in continuity with any associated internal injuries in the final autopsy report. However, of necessity, the practicalities of the dissection of the body may require that this is carried out in a somewhat disjointed fashion, particularly where there are a large number of wounds. All penetrating wounds, such as gunshot wounds and stab wounds, must be traced from their entry point through the body to their termination or exit point. All nonpenetrating injuries need to be associated with damage to the underlying tissues such as bony fractures, lacerated blood vessels and associated hemorrhage.

Special procedures

Special procedures utilized during the external examination include photography for the purposes of identification and documentation. Infrared and ultraviolet photography will enhance tattoos, bruises and patterned injuries. High-contrast black-and-white photography or computer-directed image enhancement can be used to enhance patterned injuries. Trace material can be identified with ultraviolet, laser or alternative light sources. Fingerprinting may be required for identification purposes, and is routine for all homicide victims. Where sexual assault is suspected, the collection of physical evidence includes what would normally be collected in a living victim. Insect specimens, such as fly maggots, may be useful for estimating time of death or for toxicological studies. Gunshot residues can be collected from the skin surface. Radiological examinations assist in identification, locating foreign objects such as projectiles, and documenting old and recent bony injury, which is of particular importance in suspected child abuse, when full body X-rays are required. Angiography and more sophisticated techniques such as computerized tomography (CT) and nuclear magnetic resonance (NMR) scanning may be useful in special circumstances.

For many deaths the information obtained from the history, scene of death and external examination is sufficient and no dissection of the body is required. Such a postmortem external examination of the body is distinguished from an autopsy, which always involves dissection. The purpose of continuing the examination with a dissection is to obtain further information not otherwise available. This internal examination almost always requires the opening of all three body cavities, namely the head, chest and abdomen. In some jurisdictions the opening of all three body cavities is mandatory under regulations governing the conduct of medicolegal autopsies. Even where no such rules exist, it is unwise to omit the examination of any of the three cavities. Such an incomplete autopsy fosters lingering doubts, brings into question the competence and judgment of the pathologist and may unnecessarily precipitate a reopening of the investigation. It is a particularly undesirable practice where the body is to be shipped overseas for disposal, when distance and the differences in language, culture and legal processes combine to compound misunderstandings.

Internal Examination

The internal examination of the chest and abdominal cavities is made through a large midline incision to the anterior chest and abdominal wall. This may be a simple linear incision from lower neck to pubic symphysis, a T-shaped incision from shoulder to shoulder and lower neck to symphysis, or a Y-shaped incision sparing the lower neck and upper chest. The choice of incision is dictated by the required exposure, but also by the cosmetic effect, bearing in mind the likely funerary arrangements. The chest plate is removed as a triangle of sternum and costal cartilages and the internal organs removed as a single block, as groups of organs, or as single organs, depending upon the preference and the judgment of the pathologist. When the organs are removed singly, the order of their removal may be modified according to the specific requirements of the case. All of the internal organs including the pelvic organs, comprising the bladder, internal genitalia and rectum, are removed. Using a coronal incision of the scalp, from ear to ear, the scalp is peeled forwards and back to expose the cranial vault, the top of which is cut off to allow removal of the brain and examination of the interior of the skull. The major organs are always weighed and the weights recorded. Where no special techniques are employed the method of dissection is similar to that of a 'hospital' autopsy. As well as recording evidence of trauma and natural disease, important negative observations, such as the absence of coronary artery disease, pulmonary

thromboemboli and bony fractures, are also recorded. This serves to provide valuable exclusionary information, and to document the completeness of the internal examination. The quality of documentation required of both the external and the internal examinations is that sufficient for another pathologist to reach a reliable independent interpretation of the findings.

Special techniques

Some types of trauma and complications of trauma require special autopsy techniques. In cases of suspected assault the muscles of the anterior shoulders, the anterior abdominal wall and the back are examined. If these muscles are not exposed and dissected then bruises within them may escape detection, particularly since there is often no visible injury on the skin surface. Venous air embolism is a potentially lethal condition and can cause sudden death in association with abortion, labour or penetrating injuries to the neck, such as stab wounds. The diagnosis may be suspected from the circumstances of the death but ultimately rests upon the observation of air in the right side of the heart and great veins at the time of autopsy. This air may be seen on a chest or abdominal X-ray of the decedent. Careful exposure of the inferior vena cava and the opening of the heart under water at the very beginning of the dissection of the body is essential if it is to be visualized. If not considered at the outset, air embolism is easily missed and the evidence destroyed by the routine dissection procedure. Deaths related to abortion, labour or violent sexual assaults require the removal of the internal genitalia in continuity with the external genitalia and perineum.

The most important special dissection technique used in forensic pathology is the dissection of the neck. Pathological findings in the neck are of crucial importance in deaths from hanging, ligature strangulation, manual strangulation and impacts to the head and neck. In these cases the physical evidence of injury may be very little when contrasted with the fatal outcome. The forensic dissection technique allows for examination of the neck structures *in situ*, layer by layer, in a bloodless field, i.e. after draining the neck of blood. In this way the creation of false hemorrhages as a result of the dissection technique is avoided, and even tiny areas of true hemorrhage can be identified. Failure to use this technique may result in the production of false hemorrhages, which are then erroneously interpreted as evidence of trauma in life.

Ancillary investigations

Ancillary investigations, which support the medicolegal autopsy, include a potentially large range of hospital laboratory tests and forensic laboratory examinations. In practice, the most important of these is toxicology because of the prevalence of prescribed and illicit drug usage, which may represent a cause or contributory factor in a death. Approximately one-third of all unnatural deaths have evidence of recent alcohol ingestion, and consequently alcohol analysis is the most common toxicological investigation. Specimens for toxicological analysis should include peripheral blood (and not blood from the torso), vitreous fluid, urine and liver, as well as stomach contents where there is suspicion of recent ingestion. The original volume of the gastric contents should be recorded so that the drug concentration found in a sample on toxicological analysis can be used to calculate the mass of drug in the stomach. Stomach contents may also be relevant to estimation of time of death. Biochemical studies can be performed on blood, urine, cerebrospinal fluid and vitreous fluid, although potential testing is more limited than in the living because of interfering postmortem biochemical changes. Samples from all major organs should be preserved in formaldehyde for possible histological examination, the extent of which is at the later discretion of the pathologist. Histological examination may identify disease states not apparent during autopsy dissection and assist in aging injuries and natural diseases, such as myocardial infarction. Samples for microbiological analysis are taken in accord with hospital autopsy practice and are of particular importance in the forensic setting of a sudden infant death, when occult infection is a common cause. Increasingly, DNA samples are obtained at autopsy for purposes of identification, archiving or the deletion of decedents from DNA criminal databases.

Report and Certification

The endproduct of the autopsy is the written and signed report submitted to the legal authority instructing the examination. The pathologist should retain a duplicate signed original. The report should contain all relevant administrative information, such as the time, date and place of examination, the authorizing legal authority and the method of identification of the body, as well as the examination findings, results of ancillary investigations and chain-of-custody details. The report should always include a section offering an opinion on the interpretation of the autopsy findings in the light of the other available investigative information. Only the pathologist is in a position to advise the legal authorities on the significance of the autopsy findings, and to fail to do so through a permanent written record defeats the purpose of the autopsy. In some jurisdictions a pathologist may omit this opinion in a contentious

case, such as a death in custody or an alleged extra-judicial killing, because of fear of reprisals or because of political sympathy with the state authorities. The absence of a full, fair and frank opinion at the conclusion of an autopsy report is always a matter of concern, but particularly in these cases. In those countries where the practice of offering no interpretative opinion in such contentious deaths has become institutionalized, it serves as a marker for the corruption and lack of independence of the medicolegal death investigative system.

In addition to the autopsy report, the death certificate provides a documentary record of the death. The person who signs the death certificate records not only the cause of death but also the identity of the person and the time and place of death. In some death investigative systems this responsibility rests with the legal authority, such as the coroner, but in other jurisdictions it may lie with the autopsy pathologist. Formulation of the cause of death on the death certificate is in accordance with international rules approved by the World Health Organization. These rules were established primarily to allow for the classification and coding of deaths with a view to acquiring community-based statistics for health monitoring and planning purposes. A cause of death formulated according to these rules cannot express the complexities which may be of interest in legal proceedings, so that, in the courtroom, the certified cause of death represents only the starting point for comment.

See also: **Causes of Death**: Scene of Death. **Pathology**: Post-mortem Interval.

Further Reading

Randall BB, Fierro MF and Froede RC (1998) Practice guideline for forensic pathology. *Archives of Pathology and Laboratory Medicine* **122**:1056–1064.

Histopathology

M A Green, University of Sheffield, Sheffield, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0588

Histology (the microscopical examination of tissues) forms an integral part of modern pathological investigation. In 1991, the British Royal College of Pathologists issued guidelines for the performance and reporting of hospital autopsies in which it was stated that 'an autopsy is incomplete without histological examination of all major organs'. The same is true of the forensic autopsy. Histology is essential to confirm the nature of any natural disease noted during the naked eye examination, to identify those not visible to the naked eye (e.g. myocarditis) and is also applied in an endeavor to age injuries and natural disease processes such as myocardial infarction.

Over the last half century, advances in histopathological techniques such as enzyme histochemistry have been applied to forensic studies, but with very limited success. Like temperature recordings as a method for estimating postmortem interval, changes in tissues, whether due to disease or injury, are subject to numerous variables, and any conclusion drawn from such studies must be hedged about with qualifications.

General histopathology has made great strides in recent years. Immunocytochemistry, tumor markers, aspiration techniques and flow cytometry are routinely used in diagnosis. They have proved of limited value to the forensic pathologist. Even so, this is no excuse for limiting the amount or failing to take histology as a routine. Formalin-fixed organs and tissues can be kept for weeks or months, and once appropriately embedded in paraffin wax, they can be stored more or less permanently. Ideally, blocks from the major organs, heart, lung, liver, kidney, spleen and brain, should always be taken, and hematoxylin–eosin (H & E) stained sections should be examined. Unfortunately, in many jurisdictions such comprehensive practise is limited by both cost and time. In the investigation of unexpected infant deaths, the collection of tissue should be even wider. The British Confidential Enquiry into Sudden Death in Infancy (CESDI) lays down a protocol that requires the examination of more than 40 blocks selected from specific sites within a wide range of organs.

The selection and preparation of blocks in itself poses problems. Blocks taken 'at the table' from the brain are of little value; the brain should be fixed for at least 8 weeks, and preferably 12 weeks, before 'blocking out' is undertaken. Similarly, the hearts of middle-aged or elderly subjects should not only be formalin fixed, but decalcified if an adequate examination of the coronary artery system is to be made. Every forensic pathologist should have access to a laboratory capable of cutting frozen sections, providing sledge microtome facilities for cutting large tissue blocks, and employing technicians skilled in the selection and use of special stains. Furthermore, pathologists should conform to standard site selection protocols, for example for the examination of the brain and heart, so that their work can be subjected

to appropriate peer review for educational or legal reasons.

Histological examination of the heart is essential in the investigation of all transportation accidents. Myocarditis, although uncommon, may be identified as a possible cause of pilot/driver failure. Likewise, viral myocarditis may be found in both infants, and particularly in older children, where a thorough gross autopsy has failed to demonstrate a cause of death.

The histological examination of the lungs may reveal acute infection, even in those cases where microbiological studies are negative, evidence of long-standing industrial lung disease, or deposits of starch, talc and other refractile materials in intravenous drug users (**Fig. 1**). Other chronic illnesses predisposing to sudden death, such as asthma, are often recognized microscopically. Widespread inhalation of vomit, as opposed to agonal spillage, can be demonstrated. It should always be remembered that aspiration *per se* is not an adequate cause of death; the underlying reasons must be sought.

In recent years, consequent upon the work of Meadow and others, the importance of Munchausen syndrome by proxy has been more fully appreciated. Children may be harmed by their carers, who are seeking attention for themselves rather than intending to kill. Such acts, for example deliberate airway obstruction, may have fatal results.

Suffocation *cannot* be proved conclusively by histology. Alveolar distention and rupture may be a consequence of attempts at resuscitation. There may be some bleeding into the alveoli in natural sudden infant deaths (SIDS, cot death). However, the presence of large amounts of fresh bleeding should arouse suspicion, and if iron-laden macrophages (siderophages) are seen both in the alveoli and the supporting tissue this should be regarded as a marker of possible previous suffocative episodes. Hemosiderin deposits associated with delivery are thought to clear within 6–8 weeks; natural diseases such as primary idiopathic pulmonary hemosiderosis are rare, particularly in children under the age of 1 year. Bleeding dyscrasias show evidence of bleeding in other organs as well. Areas of pulmonary contusion, or even frank infarction, may underlie fractures of the ribs in physical child abuse (**Fig. 2**).

In the head-injured infant, as in those in whom head injury is suspected, histological examination of the eyes and brain is mandatory. The distribution and pattern of retinal hemorrhage seen in shaking, with or without impact, is typical, although not absolutely diagnostic (**Fig. 3**). Its severity correlates well with the extent of intracranial injury and brain damage.

The brain may show very little change when examined with the naked eye, but properly selected and processed sections may show lesions ranging from frank tears of tissue and subcortical hemorrhages to diffuse axonal injury, and the release of β-amylase precursor protein. Evidence of previous trauma or episodes of hypoxia may also be found.

Any subdural clot should be measured, a section examined microscopically, and any overlying neomembranes subjected to microscopy. It has been recognized for many years that the thickness of such membranes and the degree of organization of the underlying clot may give an indication of the age of the injury. However, these are by no means reliable guides. In many cases the injury's admitted timing bears no relation to its age assessed 'blind' on its histological appearance. Subdural membranes (and where appropriate, brain sections) should be stained for iron deposits with Perls' Prussian blue reagent. The presence of hemosiderin indicates previous injury. Eye sections in such cases should also be stained for iron, although the number of positive findings is considerably lower.

Aging of injuries, both in adults and children, by microscopy is notoriously difficult. The most that one can usually say is a broad statement such as 'the degree of organization is consistent with infliction hours rather than days previously'. Attempts to apply elegant histochemical techniques to aging of injuries have proved too expensive and unreliable for use in routine practice. Occasionally, histology differentiates between true injury and innocent lesions, such as the 'Mongolian blue spot' frequently present in the skin over the lower back in infants, as pronounced and obstinate lividity. It is also of value in the recognition

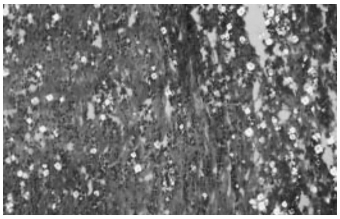


Figure 1 (see color plate 35) Intravenous drug abuse: starch granules and other contaminants at the injection site. H & E; original magnification ×200.

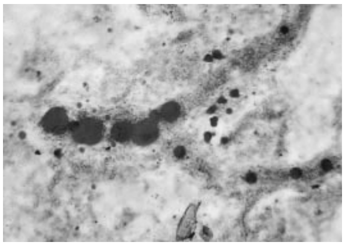


Figure 2 (see color plate 36) Pulmonary fat embolism following multiple fractures. Oil red O; original magnification × 250.

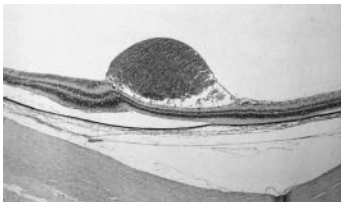


Figure 3 Preretinal hemorrhage in a 4-month-old child with severe acute brain injury. H & E; original magnification ×200.

of electrical burns and the differentiation between antemortem and postmortem thermal injury by the presence of vital reaction in the wound margins.

Histopathology is essential in the investigation of maternal death. Unless lung sections are stained appropriately, amniotic fluid embolism may be overlooked. Disseminated intravascular coagulation may also only be recognizable through the microscope after a stain for fibrin, such as Lendrum's martius scarlet blue, has been applied. In perioperative deaths, the recognition of preexisting illness may save the surgical team from accusation of negligence – or even incompetence.

A common problem for the forensic pathologist is sudden death during or shortly after an altercation, particularly when no blows appear to have been exchanged. Careful search for signs of early myocardial ischemia should be made, using phosphotungstic acid – hematoxylin or trichrome stains. Again, a few years ago there was a fashion for using enzyme stains such as TTC and toluidine blue, but these, like so many others, work within such a broad time scale that little may, with confidence, be adduced from their use.

In cases of sudden collapse during or shortly after an assault, a posterior fossa traumatic subarachnoid hemorrhage may be present. These usually result from a tear of a vertebral artery, which may be within the skull or in the high cervical spine. Not all such tears are demonstrable by flushing or radiological techniques. Step sections throughout the course of the suspect vessel may well show disruption and hemorrhage within the wall, even if the exact site of the tear cannot be demonstrated.

Histological examination is essential in the assessment of the severity of industrial lung disease. Asbestos fiber counts, carried out by those with a special interest in the techniques, are imperative where claims for compensation for lung cancer or mesothelioma are being pursued. In suspected drowning, less reliance than formerly is placed upon the presence of diatoms in lung, liver and bone, but if they are present in large number, and identical to species collected at the locus, they have some diagnostic value.

Finally, the characteristic appearance of the lungs in intravenous drug abusers has already been alluded to. Ongoing work in the UK suggests that cannabis use may be associated with structural lung damage; prevalence of hepatitis C (and no doubt other strains) renders microscopy of drug users' livers (as well as serology) mandatory.

See also: **Pathology**: Postmortem Changes; Postmortem Interval.

Further Reading

Beecroft DMO and Lockett BK (1997) Intra-alveolar siderophages in sudden infant death: a marker for previously imposed suffocation. *Pathology* **29**:60–63.

Department of Health (1996) *National Advisory Body for Confidential Enquiry into Still Birth and Deaths in Infancy, 3rd Annual Report*. London: DoH.

DiMaio DJ and DiMaio VJM (1989) *Forensic Pathology*. New York: Elsevier.

Dorandeu A, Perie G, Jokan H (1999) Histological demonstration of haemosiderin deposits in lungs and liver from victims of chronic physical child abuse. *International Journal of Legal Medicine* **112**:280–286.

Green MA, Lieberman G, Milroy CM and Parsons MA (1996) Ocular and cerebral trauma in non-accidental injury in infancy: underlying mechanisms and implications for paediatric practice. *British Journal of Ophthalmology* **80**:282–287.

Janssen W (1984) *Forensic Histopathology*. Berlin: Springer.

Knight B (1996) *Forensic Pathology*, 2nd edn. London: Edward Arnold.

Meadow SR (1999) Unnatural sudden infant death. *Archives of Disease in Childhood* **80**:7–14.

Poets CF, Southall DP, Plunkett MJ (1994) Recent developments in research into sudden infant death. *Thorax* **49**:196–197.

Taylor D (ed.) (1999) Child abuse and the eye. *Eye* **13**:3–10.

Yukawa N, Carter N, Rutty G and Green MA (1999) Intra-alveolar haemorrhage in sudden infant death syndrome: a cause for concern. *Journal of Clinical Pathology* **52**:581–587.

Postmortem Changes

M A Green, University of Sheffield, Sheffield, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0586

Introduction

The clinical diagnosis of death is beyond the scope of this article. 'Death', as generally understood, is the extinction of personality, accompanied by the loss of

vital signs such as heart beat, reflexes and respiration. This is known as 'somatic death'. Thereafter, cellular death follows over a period of time. The survival of individual organs and cells, appropriately handled, provides the basis for organ transplantation.

So-called suspended animation is mentioned here only briefly. The clinical diagnosis of death may be difficult in the elderly, the myxedematous, those under treatment with phenothiazines or under the influence of sedative and narcotic drugs. Voluntary suspended animation has been described amongst Indian holy men. Edgar Allen Poe gives a suitably gruesome account of Monsieur Valdemar's demise in his *Tales of Mystery and Imagination*.

Immediate Changes after Somatic Death

Within minutes of cessation of the heart beat, death, as understood by the laity, is usually obvious. The body becomes flaccid, the muscles lose their tone and the corneal and light reflexes disappear. The globes of the eyes become soft due to falling intraocular pressure. Upon ophthalmoscopic examination, the columns of red cells within the retinal veins break up, so-called 'trucking', and gradually tumble over the margins of the fovea centralis. This phenomenon may be observed for several hours after death, provided the cornea is moistened with water or saline to minimize the effects of clouding.

The pupillary muscles respond to mydriatics and to miotics for some hours after death. Abortive attempts were made to use these phenomena as a means of estimating postmortem interval, but proved unreliable. Similarly, voluntary muscles, particularly those of the face and limbs, may respond to electrical stimuli or direct tapping with a reflex hammer during the immediate postmortem period, but these techniques have also proved of little value.

Hypostasis (Postmortem Lividity, Sugillation)

This is one of the earliest signs of death. Blood, under the influence of gravity, begins to settle in the vessels of the dependent parts of the body within an hour or so. It may appear even earlier in the obese and 'full blooded'. Indeed, it was said to occur before death in certain wasting or dehydrating diseases such as cholera, but this author has no first hand knowledge of the phenomenon.

Pressure of even a mild degree prevents hypostasis developing. Thus a person dying on the back shows areas of pale 'contact flattening' over the shoulder blades, elbows, buttocks, thighs and calves. Hypostasis becomes fully developed within about 6 h and 'fixed' within a few hours thereafter. Thus, if a body is moved for innocent or malign reasons, the primary as opposed to secondary hypostasis, which is often paler in color, will remain and will reveal that such interference has occurred.

Hypostasis normally has a cyanotic or violaceous hue. Vanezis and his coworkers have attempted to use electrocolorimetric methods to determine the intensity of color and thus assess postmortem interval. This, like other attempts to determine postmortem interval, has proved unreliable.

In a naked body exposed to air, transcutaneous oxygenation may impart a pinker hue. This is also well seen in bodies exposed to cold, be it hypothermia or deliberate refrigeration. The marked 'pinking' may lead to a mistaken diagnosis of death by carbon monoxide or cyanide poisoning. Certain poisons which cause methemoglobinemia, for example sodium chlorate, nitrates or aniline derivatives, can produce a color ranging from slate gray to dark chocolate brown.

After about 12 h hypostasis is very firmly fixed (**Fig. 1**), but may be cleared by firm massage of the affected part of the skin with a moist sponge. Underlying bruises may then be revealed. If doubt persists as to the authenticity of a purple mark, it should be incised. In a genuine bruise the blood lies free within the subcutaneous muscle and fat. The microscopic examination of tissue from a suspect area should remove any remaining doubt.

In postural asphyxia or in cases where death has occurred with the deceased's head in a dependent position (for example an intoxicated person who has fallen half out of bed), the hypostasis may be so intense that capillaries may rupture, producing petechial type hemorrhages of varying size. However, these usually extend down below the root of the neck and on to the chest in an 'open neck shirt' distribution, so the distinction from throttling or ligature strangulation is usually easy. Hemorrhagic spots may also be

Figure 1 (see color plate 37) Hypostasis in a body that has been lying on the back for 12 h.

seen in the legs of those victims of hanging who have been left suspended for several hours.

Rigor Mortis

The flaccidity which immediately succeeds death is soon followed by the onset of progressive stiffening of the muscles. In general terms, it appears first in the small muscles of the face (hence the need to close the eyes and bind the jaw of the newly dead as part of last offices). Larger muscle groups are then affected progressively, and rigor is usually complete in 4–6 h. It may become so intense that considerable force needs to be applied to break it down and compose the body in the anatomical position. Indeed, fractures of the neck of the femur and the upper humerus may be produced by such clumsy endeavors. Similarly, postmortem fractures of the cervical spine are not uncommon. Rigor usually begins to pass off in about 18 h, and the body returns to its flaccid state in 36 h or so. The process occurs in reverse order, and residual stiffness of the fingers and eyelids often persists for the longest period of time.

In the past, great emphasis was placed upon the value of rigor in assessment of postmortem interval. This has resulted in miscarriages of justice. It cannot be too strongly emphasized that numerous factors, both intrinsic and extraneous, affect the development of rigor. Extremes of heat and cold both produce rapid stiffening, but in hot conditions it passes off quickly due to the effect of putrefaction. On the other hand, stiffening persists until such time as the body is brought into a temperate environment; it then passes off quickly.

Those who die with a raised body temperature, be it due to illness or violent exertion, may show a rapid onset and loss of rigor. Conversely, those who die of inanition, wasting disease and cachexia develop only weak and transient rigor, the onset of which may be delayed for 12 h or more.

The biochemistry of rigor mortis is still not fully understood. Its onset is accelerated by fatigue and high lactic acid levels in muscle. Muscle proteins include myosin and actin, which form a contractile compound, actomyosin, under the influence of the enzyme adenosine triphosphate (ATP). After death ATP is progressively destroyed, and the actomyosin forms a firm gel until, under the influence of putrefaction, protein disintegration and progressive loss of rigor begins.

Instantaneous rigor (cadaveric spasm)

This is rare. This author has encountered a bare handful of cases. It is usually associated with violent death, for example suicide by handgun or cut throat, and may be localized, with the weapon so firmly gripped in the hand that its removal is difficult. Even more rarely, it may be generalized, with the whole body fixed in the position in which death occurred. For example, a suicidal cut-throat victim may be found sitting bolt upright on a stool with the razor blade clutched between finger and thumb (**Fig. 2**). Perpetrators of homicide sometimes attempt to simulate suicide by placing the weapon in the victim's hand and closing the fingers around it. Under those circumstances, the rigor is more readily broken down, and the distinction is easily made. Furthermore, the weapon is often incorrectly positioned, making the stratagem even more obvious.

Late Postmortem Changes

Putrefaction (**Fig. 3**) is the final and incontrovertible sign of death. The speed of its onset and its varying pathways are greatly affected by the environment. This description is therefore restricted to events as they usually occur in temperate conditions (i.e. 20–40°C). In cold conditions putrefaction may be delayed, but not completely averted. At high temperatures, even only relatively slightly above 40°C, it may be rapid indeed.

Putrefaction is the consequence of bacterial and enzyme activity. Fungi and the depredations of insects and animals also have their part to play. The most dominant factor in the development of putrefaction is the spread of bacteria from the intestine. Thus the earliest sign of putrefaction is greenish discoloration in the right lower quadrant of the abdomen overlying the cecum. This discoloration, due to the breakdown of blood to methemoglobin and sulfhemoglobin, rapidly spreads to involve the whole of the abdomen and chest. Within 4–7 days the surface veins of the limbs also become prominent and discolored (marbling) and the trunk (and external genitalia in the male) begin to swell. The rise in intrathoracic and intraabdominal pressures results in the expulsion of fluid

Figure 2 Suicide: cadaveric spasm.

Figure 3 (see color plate 38) Putrefaction: death in a centrally heated flat 4 days previously.

from the nose and the mouth. These discharges are known to funeral directors as 'lung purges' and 'stomach purges', respectively. Feces may also be expelled from the anus. The fluid at the nose and mouth is often so heavily blood-stained that some form of assault might be thought to have taken place. By this time, the body has a foul odor due to the production of compounds such as indole and skatole.

In the next stage the gas formation extends to the head and limbs. The skin acquires a crepitant 'bubbly' texture, and blisters filled with pinkish fluid appear. These may be readily distinguished from antemortem burns and scalds by the lack of pink 'vital reaction' at their margin. The superficial layers of skin begin to peel away even upon the lightest touch (so-called 'skin slip'). Body fluids soak into the surrounding bedding, floor coverings, etc. – or even into the room below.

The onset and progress of putrefaction are hastened by pyrexia of whatever cause, infection and by any perforation of the gut. It is said that poisoning by arsenic or cyanide delays the onset of putrefaction. Certainly the only case of homicidal arsenic poisoning seen by this author was remarkably well preserved upon exhumation several weeks after death and the initial autopsy.

The internal organs putrefy at different rates. The brain rapidly liquefies. The liver first becomes foamy owing to the production of gas, then semiliquid. The spleen liquefies within a matter of hours. The pancreas and kidneys show early microscopic change, even though their general configuration is preserved. The heart and lungs may remain recognizable and often examinable by the naked eye for weeks or even months after death. The nongravid uterus and the prostate gland may still be identified in remains that are otherwise skeletonized.

Adipocere and Mummification

Traditionally these processes, in which putrefaction is arrested due to changes in skin and fat, are described separately, but they can occur together in different parts of the same body given the appropriate conditions.

Adipocere is a spontaneous form of partial preservation by the hydrolysis or saponification of body fat. It is therefore more commonly seen and is more prominent in women and in young, well-nourished children. The hydrolysed fat is at first yellowish-brown in color and of firm consistency; as time passes, it becomes white and friable. It has an offensive, pervasive and persistent odor. When ignited, it burns with a smoky yellow lambent flame.

Its formation is dependent upon the presence of adequate moisture (for example, immersion in water or burial in a bog) (**Fig. 4**). Acidity is also an important factor. Bodies buried in peat bogs, which are highly acid, or interred in coffins with oak shavings rich in tannin, may be preserved for hundreds, or even thousands, of years.

Mummification is uncommon in temperate climates. The drying out and hardening of the tissues associated with this condition require temperatures far higher than those normally achieved. This author has seen mummified fetuses (and adults) recovered from lofts in the days before fibreglass roofing insulation became commonplace, and has also encountered one case where a burglar became trapped within the ventilation duct of a large office block, and was discovered 7 years later (**Fig. 5**).

The skin becomes dry, brittle and shrunken. It is usually of dark brown colour. The tips of the fingers and toes may be lost. The hair, irrespective of its colour in life, is usually foxy red. The internal organs are also dry and shrunken, but may be sufficiently well preserved for a naked eye diagnosis to be made. Microscopy is unrewarding, both in mummified and adipocerous bodies.

Figure 4 Adipocere in a body immersed in fresh water for over a year.

Figure 5 Mummification: trapped in a heating duct for 7 years.

Postmortem Entomology and Other Artifacts

The common bluebottle or blowfly lays its eggs when the skin temperature falls below 27°C. The eggs rapidly hatch into maggots, which have a voracious appetite and invade all parts of the body, including the cranium, thorax and abdomen. Beetles lay their eggs on older bodies. Bodies immersed in water are not immune to insect activity. Caddis flies, water beetles, etc., in common with all insects, will 'nibble' at exposed surfaces, producing areas of apparent abrasion.

Animals such as rats and mice attack exposed parts of bodies. Foxes will also eat carrion. Even domestic dogs and cats will turn to their owners once the contents of the biscuit bowl have been exhausted. Fungi attack the skin even in embalmed and/or refrigerated bodies. When wiped away, the underlying exposed areas of skin are often discolored and roughened; these changes may lead to the mistaken diagnosis of bruises or abrasions sustained during life. Occasionally, a clumsily handled body may be damaged while being lifted into the coffin or wheeled through the corridors of a mortuary, producing abrasions upon the skin, usually over bony prominences. These abrasions are readily distinguished from antemortem injury. They are light brown in colour and rapidly become hard, shiny and slightly depressed. There is no vital reaction around them, and they show no microscopic changes of injury.

See also: **Causes of Death**: Postmortem Changes. **Drugs of Abuse**: Postmortem Blood. **Toxicology**: Methods of Analysis – Post Mortem. **Pathology**: Histopathology. **Time Since Death**.

Further Reading

Bate-Smith E and Bendall J (1949) Factors determining the time course of rigor mortis. *Journal of Physiology* **110**:47.

Erzinclioglu Z (1989) Entomology, zoology and forensic science. *Forensic Science International* **43**:209–211.

Kevorkian J (1961) The fundus oculi as a post mortem clock. *Journal of Forensic Sciences* **6**:261.

Knight B (1979) The putrefied body. *BMJ* **ii**:1300–1301.

Knight B (1994) *Forensic Pathology*, 2nd edn. London: Edward Arnold.

Poe EA (1957) The facts in the case of M. Valdemar. In: *Tales of Mystery and Imagination*, pp 287–294. London: Folio Society.

Polson CJ, Gee DJ and Knight B (1985) *The Essentials of Forensic Medicine*, 4th edn. Oxford: Pergamon Press.

Szent-Gyorgi A (1947) *Chemistry of Muscular Contraction*. New York: Academic Press.

Vanezis P and Trujillo J (1996) Evaluation of hypostasis using a colour measuring system and its application to the assessment of post mortem interval (time of death). *Forensic Science International* **78**:19–28.

Wroblewski B and Ellis M (1970) Eye changes after death. *British Journal of Surgery* **57**:69–72.

Postmortem Interval

D J Pounder, University of Dundee, Dundee, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0587

Introduction

Evidence of the time elapsed since death, the postmortem interval, may come from three sources: (1) the body of the deceased; (2) the environment in the vicinity of the body; and (3) information on the deceased's habits, movements, and day-to-day activities. All three sources of evidence (corporal, environmental and anamnestic) should be explored and assessed before offering an opinion on when death occurred. The longer the postmortem interval then the less accurate is the estimate of it based upon corporal changes. As a consequence, the longer the postmortem interval then the more likely it is that associated or environmental evidence will provide the most reliable estimates of the time elapsed.

No problem in forensic medicine has been investigated as thoroughly as the determination of the postmortem interval on the basis of postmortem changes to the body. Many physicochemical changes begin to take place in the body immediately or shortly after death and progress in a fairly orderly fashion until the body disintegrates. Each change progresses at its own rate which, unfortunately, is strongly influenced by

Figure 5 Mummification: trapped in a heating duct for 7 years.

Postmortem Entomology and Other Artifacts

The common bluebottle or blowfly lays its eggs when the skin temperature falls below 27°C. The eggs rapidly hatch into maggots, which have a voracious appetite and invade all parts of the body, including the cranium, thorax and abdomen. Beetles lay their eggs on older bodies. Bodies immersed in water are not immune to insect activity. Caddis flies, water beetles, etc., in common with all insects, will 'nibble' at exposed surfaces, producing areas of apparent abrasion.

Animals such as rats and mice attack exposed parts of bodies. Foxes will also eat carrion. Even domestic dogs and cats will turn to their owners once the contents of the biscuit bowl have been exhausted. Fungi attack the skin even in embalmed and/or refrigerated bodies. When wiped away, the underlying exposed areas of skin are often discolored and roughened; these changes may lead to the mistaken diagnosis of bruises or abrasions sustained during life. Occasionally, a clumsily handled body may be damaged while being lifted into the coffin or wheeled through the corridors of a mortuary, producing abrasions upon the skin, usually over bony prominences. These abrasions are readily distinguished from antemortem injury. They are light brown in colour and rapidly become hard, shiny and slightly depressed. There is no vital reaction around them, and they show no microscopic changes of injury.

See also: **Causes of Death**: Postmortem Changes. **Drugs of Abuse**: Postmortem Blood. **Toxicology**: Methods of Analysis – Post Mortem. **Pathology**: Histopathology. **Time Since Death**.

Further Reading

Bate-Smith E and Bendall J (1949) Factors determining the time course of rigor mortis. *Journal of Physiology* **110**:47.

Erzinclioglu Z (1989) Entomology, zoology and forensic science. *Forensic Science International* **43**:209–211.

Kevorkian J (1961) The fundus oculi as a post mortem clock. *Journal of Forensic Sciences* **6**:261.

Knight B (1979) The putrefied body. *BMJ* **ii**:1300–1301.

Knight B (1994) *Forensic Pathology*, 2nd edn. London: Edward Arnold.

Poe EA (1957) The facts in the case of M. Valdemar. In: *Tales of Mystery and Imagination*, pp 287–294. London: Folio Society.

Polson CJ, Gee DJ and Knight B (1985) *The Essentials of Forensic Medicine*, 4th edn. Oxford: Pergamon Press.

Szent-Gyorgi A (1947) *Chemistry of Muscular Contraction*. New York: Academic Press.

Vanezis P and Trujillo J (1996) Evaluation of hypostasis using a colour measuring system and its application to the assessment of post mortem interval (time of death). *Forensic Science International* **78**:19–28.

Wroblewski B and Ellis M (1970) Eye changes after death. *British Journal of Surgery* **57**:69–72.

Postmortem Interval

D J Pounder, University of Dundee, Dundee, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0587

Introduction

Evidence of the time elapsed since death, the postmortem interval, may come from three sources: (1) the body of the deceased; (2) the environment in the vicinity of the body; and (3) information on the deceased's habits, movements, and day-to-day activities. All three sources of evidence (corporal, environmental and anamnestic) should be explored and assessed before offering an opinion on when death occurred. The longer the postmortem interval then the less accurate is the estimate of it based upon corporal changes. As a consequence, the longer the postmortem interval then the more likely it is that associated or environmental evidence will provide the most reliable estimates of the time elapsed.

No problem in forensic medicine has been investigated as thoroughly as the determination of the postmortem interval on the basis of postmortem changes to the body. Many physicochemical changes begin to take place in the body immediately or shortly after death and progress in a fairly orderly fashion until the body disintegrates. Each change progresses at its own rate which, unfortunately, is strongly influenced by

largely unpredictable endogenous and environmental factors. Consequently, using the evolution of post-mortem changes to estimate the postmortem interval is invariably difficult, and always of limited accuracy.

Body Cooling

Body cooling (algor mortis or 'the chill of death') is the most useful single indicator of the postmortem interval during the first 24 h after death. Some authorities would regard it as the only worthwhile corporal method. The use of this method is only possible in cool and temperate climates, because in tropical regions there may be a minimal fall in body temperature postmortem, and in some extreme climates, such as desert regions, the body temperature may even rise after death.

Since body heat production ceases soon after death but loss of heat continues, the body cools. After death, as during life, the human body loses heat by radiation, convection and evaporation. The fall in body temperature after death is mainly the result of radiation and convection. Evaporation may be a significant factor if the body or clothing is wet. Heat loss by conduction is not an important factor during life, but after death it may be considerable if the body is lying on a cold surface.

Newton's law of cooling states that the rate of cooling of an object is determined by the difference between the temperature of the object and the temperature of its environment. A plot of temperature against time gives an exponential curve. However, Newton's law applies to small inorganic objects and does not accurately describe the cooling of a corpse which has a large mass, an irregular shape, and is composed of tissues of different physical properties. The cooling of a human body is best represented by a sigmoid curve when temperature is plotted against time. Thus, there is an initial maintenance of body temperature which may last for some hours – the so-called 'temperature plateau' – followed by a relatively linear rate of cooling, which subsequently slows rapidly as the body approaches the environmental temperature. The postmortem temperature plateau is physically determined and is not a special feature of the dead human body. Any inert body with a low thermal conductivity has such a plateau during its first cooling phase. The postmortem temperature plateau generally lasts between 30 min and 1 h, but may persist for as long as 3 h, and some authorities claim that it may persist for as long as 5 h.

It is usually assumed that the body temperature at the time of death was normal. However, in individual cases the body temperature at death may be subnormal or markedly raised. As well as in deaths from hypothermia, the body temperature at death may be subnormal in cases of congestive cardiac failure, massive hemorrhage, and shock. The body temperature may be raised at the time of death following an intense struggle, in heat stroke, in some infections and in cases of pontine hemorrhage. The English forensic pathologist Simpson recorded a personal observation of a case of pontine hemorrhage with a temperature at death of 42.8°C (109°F). Where there is a fulminating infection, e.g. septicemia, the body temperature may continue to rise for some hours after death.

Thus the two important unknowns in assessing time of death from body temperature are: (1) the actual body temperature at the time of death; and (2) the actual length of the postmortem temperature plateau. For this reason assessment of time of death from body temperature cannot be accurate in the first 4–5 h after death, when these two unknown factors have a dominant influence. Similarly, body temperature cannot be a useful guide to time of death when the cadaveric temperature approaches that of the environment. However, in the intervening period, over the linear part of the sigmoid cooling curve, any formula which involves an averaging of the temperature decline per hour may well give a reasonably reliable approximation of the time elapsed since death. It is in this limited way that the cadaveric temperature may assist in estimating the time of death in the early postmortem period.

Unfortunately the linear rate of postmortem cooling is affected by environmental factors other than the environmental temperature and by cadaveric factors other than the body temperature at the time of death. The most important of these factors are body size, body clothing or coverings, air movement and humidity, and wetting or immersion in water.

Body size is a factor because the greater the surface area of the body relative to its mass, the more rapid will be its cooling. Consequently, the heavier the physique and the greater the obesity of the body, the slower will be the heat loss. Children lose heat more quickly because their surface area to mass ratio is much greater than that of adults. The exposed surface area of the body radiating heat to the environment will vary with the body position. If the body is supine and extended, only 80% of the total surface area effectively loses heat, and in the fetal position the proportion is only 60%.

Clothing and coverings insulate the body from the environment and therefore slow body cooling. The effect of clothing has a greater impact on corpses of low body weight. A bedspread covering may at least halve the rate of cooling. For practical purposes, only the clothing or covering of the lower trunk is relevant.

Air movement accelerates cooling by promoting convection, and even the slightest sustained air movement is significant if the body is naked, thinly clothed or wet. Cooling is more rapid in a humid rather than a dry atmosphere because moist air is a better conductor of heat. In addition, the humidity of the atmosphere will affect cooling by evaporation where the body or its clothing is wet. A cadaver cools more rapidly in water than in air because water is a far better conductor of heat. For a given environmental temperature, cooling in still water is about twice as fast as in air, and, in flowing water, about three times as fast.

Simple formulae for estimating the time of death from body temperature are now regarded as naive. The literature is replete with formulae enthusiastically recommended at first and later disavowed. The best tested and most sophisticated current method for estimating the postmortem interval from body temperature is that of the German researcher Henssge. Even so, it is acknowledged that the method may produce occasional anomalous results. It uses a nomogram based upon a complex formula, which approximates the sigmoid-shaped cooling curve. To make the estimate of postmortem interval, using this method requires (1) the body weight, (2) the measured environmental temperature and (3) the measured deep rectal temperature, and assumes a normal body temperature at death of 37.2°C. Empiric corrective factors allow for the effect of important variables such as clothing, wetting and air movement. The use of these corrective factors requires an element of personal experience. At its most accurate, this sophisticated methodology provides an estimate of the time of death within a time span of 5.6 h with 95% probability. One of the most useful aspects of the nomogram is the ease with which the effect of changes in the variables can be tested. As a result it is an educational as well as a practical investigative tool.

The assessment of body cooling is made on the basis of measurement of the body core temperature, and, postmortem, this requires a direct measurement of the intraabdominal temperature. In practice either the temperature is measured rectally, or the intrahepatic/subhepatic temperature is measured through an abdominal wall stab. Oral, aural and axillary temperatures cannot be used because after death these are not reflective of the body core temperature. For the measurement, an ordinary clinical thermometer is useless because its temperature range is too narrow, and the thermometer is too short for insertion deep into the rectum or liver. A chemical thermometer 25–30 cm (10–12 inches) long with a range from 0 to 50°C is ideal. Alternatively a thermocouple probe can be used, and this has the advantage of a digital readout or a programmable printed record.

Whether the temperature is measured via an abdominal stab or rectally is a matter of professional judgment in each case. If there is easy access to the rectum without the need to seriously disturb the position of the body, and if there is no reason to suspect sexual assault, then the temperature can be measured rectally. It may be necessary to make small slits in the clothing to gain access to the rectum, if the body is clothed and the garments cannot be pushed to one side. The chemical thermometer must be inserted about 7.5–10 cm (3–4 inches) into the rectum and read *in situ*. The alternative is to make an abdominal stab wound after displacing or slitting any overlying clothing. The stab can be made over the right lower ribs and the thermometer pushed into the substance of the liver, or alternatively a subcostal stab will allow insertion of the thermometer on to the undersurface of the liver. If a method of sequential measurements of body temperature is to be used, the thermometer should be left *in situ* during this period. Taking sequential readings is much easier with a thermocouple and an attached printout device.

The core body temperature should be recorded as early as conveniently possible at the scene of death. The prevailing environmental temperature should also be recorded and a note made of the environmental conditions at the time the body was first discovered, and any subsequent variation in these conditions. Temperature readings of the body represent data, which, if not collected, are irretrievably lost. Therefore the decision not to take such readings should always be a considered one.

Rigor Mortis

Ordinarily, death is followed immediately by total muscular relaxation, primary muscular flaccidity, succeeded in turn by generalized muscular stiffening, rigor mortis. After a variable period of time, as a result of the development of putrefaction, rigor mortis passes off spontaneously, to be followed by secondary muscular flaccidity. There is great variation in the rate of onset and the duration of rigor mortis, so that using the state of rigor mortis to estimate the postmortem interval is of very little value. In general, if the body has cooled to the environmental temperature and rigor is well developed, then death occurred more than 1 day previously and less than the time anticipated for the onset of putrefaction (see below), which is about 3–4 days in a temperate climate.

As a general rule, when the onset of rigor is rapid its duration is relatively short. The two main factors that influence the onset and duration of rigor are the

environmental temperature and the degree of muscular activity before death. Onset of rigor is accelerated and its duration shortened when the environmental temperature is high, so that putrefaction may completely displace rigor within 9–12 h of death. If the temperature is below 10°C it is exceptional for rigor mortis to develop, but if the environmental temperature is then raised, rigor mortis will develop in a normal manner. Rigor mortis is rapid in onset, and of short duration, after prolonged muscular activity, e.g. after exhaustion in battle, and following convulsions. Conversely, a late onset of rigor in many sudden deaths can be explained by the lack of muscular activity immediately prior to death.

Classically, rigor is said to develop sequentially, but this is by no means constant, symmetrical or regular. Antemortem exertion usually causes rigor to develop first in the muscles used in the activity. Otherwise, rigor is typically first apparent in the small muscles of the eyelids, lower jaw and neck, followed by the limbs. It involves first the small distal joints of the hands and feet and then the larger proximal joints of the elbows, knees and then the shoulders and hips. It is generally accepted that rigor mortis passes off in the same order in which it develops. The forcible bending of a joint against the fixation of rigor results in tearing of the muscles and the rigor is said to have been 'broken'. Provided the rigor had been fully established, it will not reappear once broken down by force. Re-establishment of rigor, albeit of lesser degree, after breaking it suggests that death occurred less than about 8 h before rigor was broken.

The intensity of rigor mortis depends upon the decedent's muscular development, and should not be confused with its degree of development. In examining a body, both the degree (complete, partial or absent joint fixation) and the distribution of rigor should be assessed, after establishing that no artifact has been introduced by previous manipulation of the body by other observers. Attempted flexion of the different joints will indicate the degree and location of rigor. Typically slight rigor can be detected within a minimum of 30 min after death but may be delayed for up to 7 h. The average time of first appearance is 3 h. It reaches a maximum, i.e. complete development, after an average 8 h, but sometimes as early as 2 h postmortem or as late as 20 h.

Livor Mortis

Lividity is a dark purple discoloration of the skin resulting from the gravitational pooling of blood in the veins and capillary beds of the dependent parts of the corpse. Synonyms include livor mortis, hypostasis, postmortem lividity, and, in the older literature, postmortem suggillations. Lividity is present in all corpses, although it may be inconspicuous in some. The medicolegal importance of lividity lies in its color, as an indicator of cause of death, and in its distribution, as an indicator of body position. The development of livor is too variable to serve as a useful indicator of the postmortem interval, but the tradition of evaluating it remains entrenched in forensic practice.

Most authorities agree that lividity attains its maximum intensity, on average, at around 12 h postmortem, but there is some variation in descriptions of when it first appears, and when it is well developed, i.e. confluent. Hypostasis begins to form immediately after death, but it may not be visible for some time. Ordinarily its earliest appearance, as dull red patches, is 20–30 min after death, but this may be delayed for up to 2, or rarely 3 h. The patches of livor then deepen, increase in intensity, and become confluent within 1–4 h postmortem, to reach a maximum extent and intensity within about 6–10 h, but sometimes as early as 3 h or as late as 16 h. Faint lividity may appear shortly before death in individuals with terminal circulatory failure. Conversely, the development of lividity may be delayed in persons with chronic anemia or massive terminal hemorrhage.

After about 10–12 h the lividity becomes 'fixed' and repositioning the body, e.g. from the prone to the supine position, will result in a dual pattern of lividity, as the primary distribution will not fade completely but a secondary distribution will develop in the newly dependent parts. The blanching of livor by thumb pressure is a simple indicator that it is not fixed. Fixation of lividity is a relative, not an absolute, phenomenon. Well-developed lividity fades very slowly and only incompletely. Fading of the primary pattern and development of a secondary pattern of lividity will be quicker and more complete if the body is moved early during the first day. However, even after a postmortem interval of 24 h, moving the body may result in a secondary pattern of lividity developing. Duality of the distibution of lividity is important because it shows that the body has been moved after death. However, it is not possible to estimate with any precision, from the dual pattern of livor, when it was that the corpse was moved.

Areas of lividity are overtaken early in the putrefactive process. The red cells are hemolyzed and the hemoglobin diffuses into the surrounding tissues, where it may undergo secondary changes such as sulfhemoglobin formation.

Putrefaction

Putrefaction is the postmortem destruction of the soft tissues of the body by the action of bacteria and

endogenous enzymes. The main changes recognizable in the tissues undergoing putrefaction are changes in color, the evolution of gases, and liquefaction. The green discoloration seen is due to sulfhemoglobin formation. The gases produced include hydrogen sulfide, methane, carbon dioxide, ammonia and hydrogen. The offensive odor is caused by some of these gases and by small quantities of mercaptans.

Bacteria are essential to putrefaction, and commensal bacteria soon invade the tissues after death. Typically, the first visible sign of putrefaction is a greenish discoloration of the skin of the anterior abdominal wall. This most commonly begins in the right iliac fossa, i.e. over the area of the cecum, where the contents of the bowel are more fluid and full of bacteria. Any antemortem bacterial infection of the body, particularly septicemia, will hasten the onset and evolution of putrefaction. Injuries to the body surface promote putrefaction by providing portals of entry for bacteria. Putrefaction is delayed in deaths from exsanguination because blood usually provides a channel for the spread of putrefactive organisms within the body. Although it tends to be more rapid in children than in adults, the onset is relatively slow in unfed newborn infants because of the lack of commensal bacteria in the gut.

Environmental temperature has a very great influence on the rate of development of putrefaction, so that rapid cooling of the body following a sudden death will markedly delay its onset. In a temperate climate, the degree of putrefaction reached after 24 h in the height of summer may require 10–14 days in the depth of winter. Putrefaction is optimal at temperatures ranging between 21 and 38°C (70 and 100°F), and is retarded when the temperature falls below 10°C (50°F) or when it exceeds 38°C (100°F). A high environmental humidity will enhance putrefaction. Heavy clothing and other coverings, by retaining body heat, will speed up putrefaction. The rate of putrefaction is influenced by body build because this affects body cooling (see above). Obese individuals putrefy more rapidly than those who are lean. Whereas warm temperatures enhance putrefaction, intense heat produces 'heat fixation' of tissues and inactivates autolytic enzymes, with a resultant delay in the onset and course of decomposition.

There is considerable variation in the time of onset and the rate of progression of putrefaction. As a result, the time taken to reach a given state of putrefaction cannot be judged with accuracy. An observer should not assert too readily that the decomposed state of a body is inconsistent with a time interval alleged. As a general rule, when the onset of putrefaction is rapid then the progress is accelerated. Under average conditions in a temperate climate the earliest putrefactive changes involving the anterior abdominal wall occur between 36 and 72 h after death. Progression to gas formation, and bloating of the body, occurs after about 1 week. The temperature of the body after death is the most important factor generally determining the rate of putrefaction. If it is maintained above 26°C (80°F) then the putrefactive changes become obvious within 24 h and gas formation is seen in about 2–3 days.

According to an old rule of thumb (Casper's dictum), 1 week of putrefaction in air is equivalent to 2 weeks in water, which is equivalent to 8 weeks buried in soil, given the same environmental temperature. After normal burial, the rate at which the body decomposes will depend to a large extent on the depth of the grave, the warmth of the soil, the efficiency of the drainage and the permeability of the coffin. The putrefactive changes are relatively rapid when contrasted with the terminal decay of the body. When the putrefactive juices have drained away and the soft tissues have shrunk, the speed of decay is appreciably reduced.

Adipocere

Saponification or adipocere formation is a modification of putrefaction characterized by the transformation of fatty tissues into a yellowish-white, greasy (but friable when dry), wax-like substance, which has a sweetish rancid odor when its formation is complete. During the early stages of its production a penetrating and very persistent ammoniacal smell is emitted. Adipocere develops as the result of hydrolysis of fat with the release of fatty acids, which, being acidic, then inhibit putrefactive bacteria. However, fat and water alone do not produce adipocere. Putrefactive organisms, of which *Clostridium welchii* is the most active, are important, and adipocere formation is facilitated by postmortem invasion of the tissues by commensal bacteria. A warm, moist, anaerobic environment thus favors adipocere formation.

Adipocere develops first in the subcutaneous tissues, most commonly involving the cheeks, breasts and buttocks. Rarely, it may involve the viscera, such as the liver. The adipocere is admixed with the mummified remains of muscles, fibrous tissues and nerves. The primary medicolegal importance of adipocere lies not in establishing the postmortem interval but rather in the preservation of the body, which aids in personal identification and the recognition of injuries.

The presence of any adipocere indicates that the postmortem interval is at least weeks and probably several months. Under ideal warm, damp conditions, adipocere may be apparent to the naked eye after 3–4 weeks. Ordinarily, this requires some months and

extensive adipocere is usually not seen before 5 or 6 months after death. Some authorities suggest that extensive changes require not less than a year after submersion, or upwards of 3 years after burial. Once formed, adipocere will ordinarily remain unchanged for years.

Mummification

Mummification is a modification of putrefaction characterized by the dehydration or dessication of the tissues. The body shrivels and is converted into a leathery or parchment-like mass of skin and tendons surrounding the bone. Mummification develops in conditions of dry heat, especially when there are air currents, e.g. in a desert. Newborn infants, being small and sterile, commonly mummify. Mummification of bodies of adults in temperate climates is unusual unless associated with forced hot-air heating in buildings or other manmade favorable conditions. The forensic importance of mummification lies primarily in the preservation of tissues, which aids in personal identification and the recognition of injuries. The time required for complete mummification of a body cannot be precisely stated, but in ideal conditions mummification may be well advanced by the end of a few weeks.

Maceration

Maceration is the aseptic autolysis of a fetus, which has died in the uterus and remained enclosed within the amniotic sac. Bacterial putrefaction plays no role in the process. The changes of maceration are only seen when a stillborn fetus has been dead for several days before delivery. Examination of the body needs to be prompt because bacterial putrefaction will begin following delivery. The body is extremely flaccid, with a flattened head and undue mobility of the skull. The limbs may be readily separated from the body. There are large, moist skin bullae, which rupture to disclose a reddish-brown surface denuded of epidermis. Skin slip discloses similar underlying discoloration. The body has a rancid odor but there is no gas formation. Establishing maceration of the fetus provides proof of a postmortem interval in the uterus, and therefore proof of stillbirth and conclusive evidence against infanticide.

Vitreous Potassium

Several researchers have studied the relationship between the potassium concentration of the vitreous humor of the eye and the postmortem interval. However, within 100 h postmortem, the 95% confidence limits of the different researchers vary from $\pm$ 9.5 h up to $\pm$40 h. Cases with possible confounding antemortem electrolyte disturbances can be excluded by eliminating all cases with a vitreous urea above an arbitrary level of 100 mg dl^{-1}. (High urea values in vitreous humor always reflect antemortem retention and are not due to postmortem changes.) Having eliminated these cases, there is a linear relationship between vitreous potassium concentration and time elapsed after death up to 120 h. However, the 95% confidence limits are $\pm$22 h, so that the method has no real practical application. There are also sampling problems, in that the potassium concentration may differ significantly between the left and right eye at the same moment.

See also: **Pathology**: Postmortem Changes.

Further Reading

Henssge C (1988) Death time estimation in case work: I. The rectal temperature time of death nomogram. *Forensic Science International* 38:209–236.

Henssge C, Knight B, Krompecher T, Madea B and Nokes L (1995) *The Estimation of the Time Since Death in the Early Post-mortem Period*. London: Edward Arnold.

Madea B, Henssge C, Honig W and Gerbracht A (1989) References for determining the time of death by potassium in vitreous humor. *Forensic Science International* 40:231–243.

Preservation of Evidence

M A Green, University of Sheffield, Sheffield, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0585

Introduction

It is now over 70 years since Locard laid down the first principle of crime scene management, ‘every contact leaves a trace’. With recent advances in fingerprint retrieval, footwear impression enhancement by electrostatic methods, and the ability to recover DNA from even a single cell, the avoidance of contamination has achieved prime importance. Indeed, cross-contamination, real or alleged, is often of more use to the defense team than is the loss of an occasional fiber or hair – although the importance of trace evidence preservation cannot be too strongly emphasized.

In many jurisdictions, crime scene management has become an increasingly specialized area of police practice. Courses and diplomas are offered in the subject and the scenes of crime officer (SOCO) now assumes many responsibilities that were at one time shared by the pathologists, forensic scientists and other crime scene specialists. However, this is not the case in every country; there are situations where the pathologist may be the only trained professional at the scene. Furthermore, the pathologist's specialist knowledge and advice are frequently sought, especially when the locus, or the nature of materials of potential evidential value are somewhat unusual. It should also be borne in mind that a senior pathologist with a heavy case load may well have been attending scenes and performing autopsies when the most senior police officer involved was still at school, and the pathologist will certainly attend more scenes every year than the average junior detective. Pathologists, therefore, cannot devolve responsibility on to others. Their errors of omission or commission may well be exposed in court, devaluing their testimony as a whole.

The pathologist's involvement in evidence preservation

The pathologist should assume responsibility in the following areas:

- selection of appropriate samples;
- proper collection of samples;
- avoidance of loss of material;
- avoidance of crosscontamination;
- selection of appropriate containers and preservatives;
- appropriate packaging;
- assurance of health and safety;
- labeling and storage;
- maintenance of continuity;
- proper disposal of contaminated material and 'sharps';
- advice to the court on production of material in court (for example, blood-stained clothing).

These considerations apply at the scene, in the mortuary and subsequently in all areas of the laboratory investigation.

Preservation of Evidence at the Scene ('The Locus')

On arrival at the scene the pathologist should introduce himself to the senior investigating officer (SIO) and to the SOCO. Before entering the building or the cordoned-off area, a plan of action and division of responsibility should be agreed upon.

The designated approach route should be identified and rigidly adhered to. Is the pathologist to be responsible for 'taping' clothing and exposed areas of skin, or will a scientist/SOCO fulfill this role? To whom will the pathologist hand each sample, and who will assume responsibility for its recording and labeling? How many persons will actually approach the body closely? Forward planning is essential; the one thing one should not do at a crime scene is 'make it up as you go along'!

Appropriate full protective clothing should be worn. This should comprise overalls, rubber boots or overshoes, gloves and head covering. This author regards facemasks as an ineffective nuisance; they provide little protection against infection or odor, they are uncomfortable, and people tend to fiddle with them – a dangerous practice in these days of single cell DNA. Overalls without pockets are to be preferred; this avoids the temptation to take sweet wrappers, soiled tissues and other sources of contamination to and from the scene. The pathologist can always persuade the exhibits officer or SOCO to hold a clipboard and pen. It should be remembered that the handheld tape recorder has no place in these circumstances, certainly not in UK jurisdictions. The only records of evidence acceptable in UK courts are contemporaneous notes made at the time. Furthermore, tape recorders are unreliable. They are subject to damage by dropping or bad weather. If it is decided to use a tape, it can be used for no other purpose thereafter. It must be retained in perpetuity and certified copies must be made available to the defense.

The body should be photographed *in situ*, and a diagram drawn. It should then be decided what trace evidence and samples to collect at the scene, and what may safely be left until the body has arrived in the mortuary. Climate may well influence these decisions; a threatened downpour or a howling gale are both good reasons for choosing expedient removal.

Tapings from exposed skin and clothing should be taken. This author prefers low impact fingerprint lifting tape, but ordinary Scotch tape serves just as well. However, it is much more likely to adhere to gloved fingers. Each 'lift' is then pressed down upon a clear acetate sheet and appropriately labeled. Any suspicious stains on skin or clothing should be lightly swabbed with a cotton-wool bud, moistened with tap water, before the lift takes place. A control water-moistened swab should also be retained and submitted for examination. Blood spots may have originated from the victim or the assailant(s). Several should be sampled.

If clothing or other material, weapons, etc. need to be lifted from the body, their position should be recorded photographically; they should then be

individually labeled and bagged. Wet and soiled clothing should be taken in brown paper bags to an appropriate drying area before packeting and onward transmission to the forensic science laboratory.

Taking mouth and pharyngeal swabs before the body is disturbed prevents contamination by reflux of stomach contents, bronchial secretions or continuing oozing of blood from facial injuries. Semen may also occasionally be found in gastric contents (see below).

Swabs from the female genital organs and from the anus pose more of a problem. There are arguments for taking these at the mortuary rather than at the scene, particularly if conditions are cramped and the lighting is poor. On the contrary, there is always the risk of drainage of vaginal secretions on to the perianal area during transport, leading to incorrect assumptions of anal sex; the same problem may arise if the rectal temperature is taken before the swabs are collected. It has been successfully argued in court that semen found deep in the rectum was transferred there by the clumsy insertion of a thermometer probe; this can be avoided by collecting the high vaginal, introital and anal swabs before inserting the rectal probe or thermometer.

In cases where a sexual motive may be suspected, and the breasts or buttocks are exposed, it is wise to take swabs from the nipples and surrounding skin, the buttocks, the suprapubic area and the front of the neck, even if there are no stains visible to the naked eye. Saliva traces may be found, even if no bitemarks or suction bruises are visible. The perpetrator may have masturbated over the victim rather than having full intercourse. Those forced to commit fellatio may spit out the seminal fluid so that it is found on the clothing or even on the ground nearby.

The recording of the distribution of blood spatter and staining, and the collection of representative samples therefrom, is usually the responsibility of the SOCO or the forensic scientist. Even so, the pathologist should take some note of its presence and distribution. It may well relate to the pattern of injuries which is subsequently found. The pathologist should also remember that the body is only a part of the scene; the trained eye may spot a stain or fragment of tissue in an area of disturbance, or appreciate the significance of a discarded tablet or a foil wrapper which has been missed by the lay observer.

The body should then be prepared for transportation to the mortuary. Bags should be placed over the head, hands and feet. They should be loosely secured, preferably with adhesive tape. String or twine, too tightly applied, can create factitious 'ligature marks' and lead to confusion. Plastic bags are most commonly used, although some police forces in the United Kingdom are switching over to brown paper, which gives rise to less condensation and moisture artifacts. If the body lies at a fire scene, nylon bags prevent the evaporation of accelerants. The body should then be carefully lifted on to a polythene sheet. The ready-made zipper body bag is not always suitable. Most of them provide barely enough room for a body of average build which lies in the anatomical position and is free of rigor mortis; bodies with established rigor or those fixed in the pugilistic attitude following exposure to heat may not fit within the limited space. Even worse, the zippers have a habit of sticking, or breaking open during transit.

It is the pathologist's responsibility to supervise the lifting and the removal of the body. Rough handling, especially in elderly osteoporotic subjects, may produce fractures, particularly of the cervical spine and the necks of the femurs. Clumsy bumping of the body against walls, pieces of furniture or other projections can produce skin lesions that may occasionally cause confusion, although such postmortem abrasions are usually readily distinguished from injuries sustained during life. Strict instructions should be given that the body and its coverings must remain undisturbed until the pathologist's arrival at the mortuary. Ideally, the body and the sheeting should be left in the shell or casket so that its removal can also be supervised. Unfortunately, the funeral director or other body removal contractor may be reluctant to allow this; it may be needed urgently for another case.

Finally, the site where the body lay should be minutely examined for further trace evidence. The area should be photographed again. Before leaving the scene, the pathologist should check all the exhibits collected, ensuring that the record tallies with that of the police. He or she should sign the appropriate labels, give advice on the storage of perishable materials, such as tissue fluids and swabs, and issue appropriate warnings on health and safety (see below).

Examination at the Mortuary

The pathologist's responsibilities for preservation of evidence continue throughout the autopsy and its aftermath. Occasionally it may become necessary to collect further samples, particularly for toxicology or DNA profiling, at a later date. As at the scene, the pathologist should work as a member of the team, and adhere to a previously agreed protocol, which may be modified appropriately in special circumstances.

Samples and exhibits are collected for three principal reasons. Firstly, to establish the identity of the deceased if this cannot be determined by conventional means. Secondly, to demonstrate any form of contact,

including sexual, between the deceased and the perpetrator(s), and, finally, to establish the cause of death, or the presence of any extraneous factors that might have contributed to it, for example microscopic evidence of natural disease, or the presence of alcohol, drugs or other toxic substances.

External examination

All those present in the autopsy suite should wear full protective clothing and gloves from the outset. Only the SIO, SOCO and photographer should be 'about the body'; other spectators should be confined to a designated viewing area.

Clothing and connected property

The pathologist should assume responsibility for lifting of the body on to the autopsy table, the removal of the protective plastic wrappings and bags, and an initial close scrutiny of the clothing *in situ*. Each exhibit should be bagged separately and labeled appropriately.

Careful examination of the clothing may reveal the presence of hitherto unnoticed stains, fibers, dirt and other materials. It is also important to record the presence and direction of bloodstains and splashes before they have been obscured by the removal of the clothing. These may provide information as to the position of the assailant during the assault, and any movement of the body thereafter.

Each item of clothing should be removed intact, the pockets searched, and then bagged (wet clothing should be allowed to dry out first). Sometimes a suicide note may be found in the pocket, leading to the immediate solution of the 'crime'. Remember that drug abusers may have unsheathed used needles wrapped in a handkerchief or tissue, so proceed with caution. In infants, the presence of blood-stained mucous stains about the mouth, on the clothing or on the bedding, is particularly significant. Natural sudden infant death may be associated with slightly pink respiratory secretions. Frankly blood-stained material should raise the possibility of mechanical asphyxia, particularly if exudation continues hours after death.

In cases of stabbing, the clothing damage should be matched to the injuries on the body. Suicides, for example, tend to lift or open the clothing first and then stab the exposed area. In some shootings, a spent bullet or fragment thereof may lie in the folds of the garments, and can be readily lost if care is not taken. Furthermore, in any case of shooting, the hands and clothing should be swabbed for firearms residues before undressing is undertaken. Different police forces use different techniques and reagents for this, and so local protocol should be followed.

If any form of penetrating injury (not just gunshot wound) is present, radiographic examination should be considered. The broken tip of a knife blade or the fragment of some other weapon may be located; so might drug-laden condoms or other containers in the intestines or body orifices. A full skeletal survey is mandatory in any case of suspected physical child abuse. The limbs, head and trunk should be individually X-rayed in all the appropriate planes by a competent radiographer. The so-called 'two plate babygram' does not reveal adequate detail of subtle bony injury, such as hairline fractures of long bone shafts or 'slipped epiphyses'.

The collection of 'intimate samples' should now be undertaken. A generous handful of plucked head hair, roots and all, should be collected. Hair should also be collected, with roots, from the eyebrows, pubic region, beard and moustache. Use disposable forceps or the gloved hand. Nail scrapings should be taken with a cocktail stick, or the nails clipped with freshly cleaned scissors. These may be used for DNA profiling. The sample is also of use to the toxicologist searching for poisoning by agents such as arsenic or thallium, and for evidence of use of drugs of abuse over a period of time. Swabs from the mouth, anus, vagina and penis should be collected. Plain swabs are used. They should be stored frozen. Blood substance grouping is no longer used routinely in England and Wales; in those jurisdictions where these techniques are still in use, samples (again on plain swabs) should be stored at 4–8°C. It is wise to take a large number of swabs (at least three), particularly from the vagina and the anus. This ensures that even small amounts of material are recovered. Some pathologists and forensic scientists use unwaxed dental floss for recovery of traces of semen from between the teeth.

Swabs should be taken from any suspicious stains elsewhere on the body. Special attention should be paid to any matted or crusted pubic hairs. The axillary hair should also be examined; semen may, albeit uncommonly, be recovered from the armpits. A comb may be used to recover material from crusted areas of hair.

Any dust, grit or earth soiling of the body should be sampled, either by taping or swabbing. Some police forces are attempting to recover fingerprints and footwear impressions from skin and from clothing. It would appear that these attempts have so far met with little success in the UK. The FBI are experimenting with ultraviolet lighting. Other methods, such as iodine fuming, have been abandoned.

If the body is colonized by insect pupae or maggots, these should be recovered and preserved in an alcohol–glycerol mixture; formaldehyde is not suitable.

Internal examination

Blood and urine should be recovered under direct vision, rather than by blind probing through intact skin. Uncontaminated urine, for both toxicology and microbiology, is best obtained with a needle and syringe through the dome of the bladder. If the bladder is almost empty, it may be opened and a few milliliters recovered from its base with a syringe devoid of needle.

The choice of site of blood samples for toxicology remains controversial. More than 20 years ago it was shown that there were wide variations in barbiturate levels, depending upon sample sites. More recently, other workers have confirmed these observations using other drugs. Furthermore, they have demonstrated that diffusion of drugs from the stomach may give artificially high levels in blood from the heart or the venae cavae. It has been well known for many years that samples taken from these sites are unreliable owing to contamination from the liver, where concentrations are frequently much higher than they are in peripheral blood. Cavity blood should never be collected for any purpose; it is certain to be contaminated with intestinal organisms. The ideal sites for blood collection are the femoral or common iliac vein. Ideally, separate samples should be taken from each side, and the lower of the two values obtained should be regarded as being most acceptable for court purposes. If the abdomen is heavily contaminated, small amounts of blood can be recovered from the femoral or popliteal vessels.

Ocular vitreous humour is the fluid of choice for biochemical tests such as electrolyte, urea and glucose levels. It may also be used for toxicology in mutilated bodies, those that have been exposed to extreme heat, or are undergoing early putrefactive change. The vitreous should not be taken at the start of the autopsy in any child where head injury might be even a remote possibility. Its extraction renders the eyes useless for histological examination. Only after the scalp, skull and brain have been examined should vitreous humour be taken in these circumstances. In child deaths, cerebrospinal fluid should be taken at the commencement of the autopsy by direct vision from the cervical canal, using a pipette, after the vertebral arches have been removed. This fluid can be used for microbiology, toxicology and biochemistry.

If poisoning is suspected, samples of liver and brain should be taken and frozen in separate containers. Most modern toxicology laboratories can carry out a full range of tests on as little as 20 g of these organs. The stomach contents should be collected; the forensic scientists may wish to determine the composition of the last meal (and sometimes look for the presence of semen) before passing them on to the toxicologists. Only rarely is it necessary to retain the contents of the small bowel. Few toxicology laboratories now require the stomach itself, but in cases of poisoning by such materials as arsenic, the advice of the toxicologist should be sought before the postmortem is concluded. The pathologist should routinely take tissue samples from all major organs for microscopy. An autopsy without full histology is regarded by truly professional pathologists as incomplete. It is certainly not acceptable in the modern court process, be it civil or criminal.

Preservation of Samples

Many toxicology laboratories are happy to accept unadulterated urine, provided it has been refrigerated and delivered in a reasonable time (24–36 h). A few still prefer the addition of a preservative such as a small amount of phenyl mercuric nitrate.

Blood is more of a problem. Samples for DNA profiling should have no preservative added, but should be frozen. It is now generally agreed that measurement of blood alcohol levels is unreliable if commercial fluoride/oxalate containers are used. These are perfectly adequate for fresh venous samples from living patients which are to be analyzed within a very short time. Postmortem samples are liable to production of alcohol by microbiological action, and higher concentrations of sodium fluoride are required to inhibit this. Forensic science laboratories now recommend the addition of sodium fluoride to a concentration of at least 1% weight for volume, i.e. a generous 'knifepoint' of the crystals to a 5 ml sample bottle. 'Vacutainers' or similar prepacked venepuncture appliances are not suitable for use at autopsy. They work well for venesection in the living, but they are difficult to use when extracting blood from the cut end of an iliac vein. A plain syringe – or even an accurately placed (and clean) dessertspoon – is to be preferred.

Solid organs and stomach contents should be frozen. Vitreous humor and urine should be refrigerated at 4–8°C. Samples for microbiology are liable to yield fallacious and misleading results. Even with prior 'searing' of the surface of the organ, for example the heart or lung, with a hot scalpel blade, postmortem contamination is almost inevitable. In the vast majority of cases, a simple nutrient blood culture bottle, and plain swabs rather than those coated with or immersed in transport medium, should be used. It should be made clear to the receiving microbiology laboratory that these samples are from postmortem material. The microbiologist who issues the report

should then be able to interpret more accurately the significance of any organisms that are grown.

The pathologist's final responsibility is to check the exhibits log, sign all the labels, and ensure that all the specimens are secure from leaks and that appropriate advice on storage is given and is clearly understood. The risks of virus infections, notably *Human immunodeficiency virus* (HIV) and bloodborne hepatitis, are now higher than formerly. In any case where the deceased may have had a suspect lifestyle (for example, sexual promiscuity or intravenous drug abuse), consideration should be given to appropriate screening of a blood sample.

See also: **Alcohol**: Blood; Body Fluids. **Crime-scene Investigation and Examination**: Recording; Recovery of Human Remains; Packaging; Preservation; Contamination.

Further Reading

Eckert WG (1992) *Introduction to Forensic Sciences*, 2nd edn, pp. 97–106. New York: Elsevier.
Forensic Science Service (1994) *The Scenes of Crime Handbook*, 2nd edn. London: Forensic Science Service.
Randall B (1997) *Death Investigation; the Basics*, pp. 21–30. Tucson, AZ: Galen Press.
Weston N (1998) In: White P (ed.) *Crime Scene to Court*, pp. 33–43. London: Royal Society of Chemistry.

Victim Recovery

J Leditschke and R Ellen, Victorian Institute of Forensic Medicine, Melbourne, Australia

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0583

Introduction

The way in which a body is retrieved from a crime scene, preserved and transported can greatly affect the appearance and subsequent evaluation of any evidence or injury. Therefore, the officer in charge of the crime scene, the forensic pathologist and the body transporter must liaise to ensure the body is handled and transported in such a way that additional injuries or artifacts are not produced nor evidence lost. In addition, any evidence noted on the body at the scene should remain in its original position; for example, a ligature in a case of a suicide by hanging. The position of this evidence should be protected while the body is in transit to the mortuary. Furthermore, any precautions taken to prevent this evidence being destroyed should be documented. For example, when paper bags are placed over the hands to preserve gunshot residues, the pathologist is informed of this before the autopsy examination to ensure there is no confusion over when the paper bags were placed on the hands – before or after death.

To ensure adequate transportation, all personnel involved with the case should be trained in procedures dealing with continuity of evidence and protocols for handling suspicious deaths. In such cases a police officer should accompany the deceased to the mortuary and oversee the transfer of the body. This is crucial to protect the family, police and mortuary staff from discrepancy when handling not only evidence but also property. The way in which clothing and property is handled can become a contentious issue for the family. Personnel recovering and transporting bodies need to be aware of protocols to maintain the continuity of the property and evidence. In addition, all personnel must wear the appropriate safety attire and be trained in procedures to minimize infectious risks. Body handling also carries an element of risk with regards to physical strain. Staff need training in special lifting techniques.

This article outlines the techniques necessary for preserving the body and associated evidence during transport, including specific considerations for different types of case.

Preservation of the Body Prior to Removal

In many cases, the clothing, body and hands of the deceased should be protected from possible contamination during transportation. Often a clean plastic body pouch or plastic sheet is enough to ensure evidence is not lost. Any additional measures taken to assist with the preservation of evidence should be clearly noted, and only performed in the presence of the pathologist or homicide detective. Once the police, pathologist and crime scene examiners have completed their examination, the deceased is prepared for transportation. Often minimal preparations are needed. If the body is prostrate and no outward injuries are observed, it can be carefully lifted into a bag and on to a stretcher. In the past, funeral directors and hospital staff have 'packed and prepared' the body. This included laying it in the anatomical position with the hands and feet tied and cotton wool placed in all orifices. These procedures should not be performed in forensic cases as they may create postmortem artifacts or injury, which may be difficult to interpret.

Incinerated cases

Severely burnt or incinerated bones are often very brittle and fragile. In these cases the identification of the deceased person can be difficult. Most often antemortem dental records will be used for comparison with the dentition of the deceased person. In many instances the mandible is exposed during a fire and is severely burnt. The teeth may be lost during transportation to the mortuary. To prevent this, the head is photographed *in situ* and then wrapped in cotton wool and bubble wrap, and is supported by a plastic bag or container. Thus, if the teeth are dislodged they remain contained within the bag. In addition, after removal of the body, the ash and debris in the region below the deceased and the immediate vicinity are carefully searched and sifted to screen for teeth, small bones or other matter important to the case. The debris may be bagged and labeled according to the section of body immediately above it, and sifted at the mortuary. It is important that anything found at the scene is bagged separately and clearly labeled: for example, 'tooth found in the vicinity of body ×'. In cases where a number of bodies are found close together it should not be assumed that separate parts belong to any particular body until further examination has been carried out.

Suicides

In many cases of suicide the items contributing to the death are found on or near the body. It is important that anything attached to the body remains attached and its position preserved as much as possible. This is to ensure that the pathologist can examine the device, e.g. wires or electrical equipment, in relation to the position of the device and any injury it may have caused. This can help to rule out the involvement of any other person in the death. In addition, in more obscure cases, if the evidence, e.g. a plastic bag over the head, is removed prior to the pathologist's examination the mode of suicide is often unclear. In the event of a hanging suicide, where the deceased is suspended, the ligature should remain intact and *in situ*. The rope should be cut away from the deceased, leaving the knot intact. If resuscitation has taken place and the ligature has been removed, the section of the ligature originally around the neck of the deceased should be reconstructed, the ends tied together with string, then bagged, labeled and transported with the body. When there is evidence with legal requirements for transport, such as firearms, drugs or volatiles, the chain of custody should be documented. The police or ballistics expert experienced in safe handling procedures usually transports firearms and other dangerous weapons.

Suspicious deaths

Preservation of evidence is most important for homicide or suspicious death. Methods vary with the circumstances. The body should only be touched or moved in consultation with the homicide squad and/or forensic pathologist. Prior to removal, the body should be carefully placed on a sheet of plastic and then into a sealed body bag to insure no fibers or trace evidence is lost. In the majority of these cases it is important that the clothing and evidence remains absolutely undisturbed until examination by the forensic pathologist. For example, in cases of sexual assault the disarray of clothing provides the pathologist with an indication of such an assault, and would highlight the possibility of injury in certain areas. In addition, the clothing folds may contain evidence such as semen, hairs or fibers that would otherwise be disturbed in removal of the clothing or transport of the body. However, in certain circumstances it may be advantageous to remove clothing or evidence at the scene; for example, to preserve blood spatter evidence. The removal of clothing, property or evidence should only occur after consultation with the forensic pathologist. In such cases chain of custody is documented to maintain continuity of evidence.

Gunshot cases

When a shot is fired, gunshot residues are ejected from the weapon on to the hands and clothing of the person firing the weapon. By measuring the quantity of residues on the persons present at the scene of death it is possible to exclude some suspects and approximate the position of any other people present at the time of the event. To preserve the gunshot residues on the deceased person during transit, paper bags are placed over the hands. It is essential that paper bags be used rather than plastic as the hands may sweat in the plastic, which would alter the presence of the residues. After the removal of the deceased on to a stretcher, the immediate vicinity is searched carefully for the projectile. The body transporter must be aware of this as the projectile may be on the surface of the skin or in the clothing and it may be disturbed when the body is moved.

Decomposed cases

The decomposing body presents a difficult situation in terms of removal and transport. The body is very fragile, and often beginning to break apart. Moving a severely decomposed body can complicate this further and result in the removal of skin and/or the detachment of limbs. To overcome this, rather than lifting the body on to plastic it should be rolled on to its side, and the plastic tucked underneath. The body

can then be lifted using the plastic to hold the remains intact. Dealing with a decomposed body is unpleasant, and there are a number of associated risks. Flies and maggots are often present, as are spiders and other insects. Care should be taken to avoid being bitten or stung; however, fly spray or deodorizer should not be used on the body as it may interfere with toxicology results. Fluid-filled blisters, called bolus, may rupture and splash body transporters, and appropriate safety attire should be worn.

Skeletal remains

Many forensic anthropology textbooks describe in detail the methods involved in recovering skeletal remains. The process is slow and meticulous as the soil layers and vegetation are removed layer by layer without disturbing the skeleton. When the body is uncovered, soft brushes are used to scrape away the remaining soil around the bones, and each is examined and removed by an anthropologist. The bones are packed in paper bags, either grouped or separately depending on the case. In some circumstances the skull may need to be wrapped similarly to the procedure used with incinerated remains, to preserve the teeth. Alternatively, the teeth may be glued in place before removal of the skull. In circumstances where some tissue is still attached to the bones, it may be possible to slide a large board underneath the deceased and lift the remains with the board, keeping it intact for transportation. As with incinerated bodies, the soil around the deceased should be bagged, labeled and sifted later to search for teeth, small bones and other evidence.

Diving fatalities

The recovery of diving fatalities is one of the more difficult situations in body removal. The equipment and its settings may provide vital clues as to the cause of death; thus it should all remain intact and should accompany the deceased *in situ*. This makes for a bulky and awkward body transfer. If an expert is present, the equipment should be checked at the scene to ensure valves are turned off and settings are not accidentally altered during transit. It is crucial that the body be transported to the mortuary as quickly as possible. In diving fatalities, radiography for air embolism needs to be performed on the deceased within 4 h, otherwise the air becomes reabsorbed into the blood and the evidence of air embolism lost.

Infectious cases

Universal precautions should be used with every body removal, regardless of its known infectious state. Gloves and care with sharps should be exercised, and all bodies treated as potentially infectious. If a risk is noted, extra care should be taken; double gloves may be used and care taken to avoid contact with body fluids. To prevent splashes, a towel or absorbent material can be placed over any open wounds or over the face of the deceased person. In the elderly, or instances where there is a risk for tuberculosis, a towel should be placed over the face to prevent the escape of sputum or other fluids. In high classifications of infection, respirators should be worn when handling the body. The deceased is wrapped in plastic, sealed in a body pouch and clearly labeled as 'high risk of infection'.

Chemical contaminations

Before a chemically contaminated body is recovered and transported, the chemical is identified and assessed for the degree of hazard. This is to insure the appropriate precautions are taken to maintain a safe work environment for scene workers, body transporters and mortuary staff. In many instances, once the body is removed from the scene the level of contamination is low. If the chemical cannot be identified it should be treated as highly toxic and maximum protection worn by workers. In extreme cases the body is decontaminated before removal. Generally, the fire brigade controls the scene. Only personnel trained in decontamination procedures should enter the scene.

It is recommended that the body be photographed *in situ*, then the clothing and property bagged and labeled, with scene workers wearing self-contained breathing apparatus. The deceased is then repeatedly hosed. In any case where there is suspicion of chemical contamination, a safety officer should be assigned to monitor continually the level of toxic fumes and insure the safety of the staff is adequate. The type of respirator necessary for transportation to the mortuary depends on the chemical. The fire brigade or the scene controller can be consulted. Each situation needs to be separately and fully assessed.

Body Lifting and Moving

The different methods of moving a body have been described above. The aim is to prevent any further injury to the deceased person and also to the personnel moving the body. It is recommended that, ideally, four people should be involved in moving a body of average size/weight from the scene to a body bag and then on to a stretcher. This allows for each person to hold a limb of the deceased or a corner of the body bag. Standard lifting techniques should be used; that is, bending the knees and keep the back straight.

With obese bodies it may be easier to roll the deceased one way, placing a the body bag beneath, and then rolling the body the other way, pulling the body bag across. This technique can also be used for cases in which decomposition has occurred. It is important that the body is not dragged, as this will cause additional injuries.

Body bags

The type of body bag will depend on the budget of the organization. Deluxe body bags can cost anything up to 10 times the standard price. In large organizations this is just not practicable; however, there is still a minimum set of requirements all body bags should possess. These include:

- tough waterproof material;
- strong zips with covers to prevent leakage;
- generous size to allow for bodies with rigor mortis and extensive decomposition;
- handles.

Property and clothing

The handling of property and clothing is an area where emergency personnel and mortuary staff are often criticized. It is therefore vital that all property and clothing is dealt with according to strict protocols.

In routine cases it is preferred that all clothing and property is left at the home. In such instances the clothing is removed and clearly recorded. Property is also clearly recorded on a property sheet. Generic terms should be used at all times; for example, a gold ring with diamonds should be described as a 'gold colored' ring with clear stones. It is also best that the removal of clothing and property is done in the presence of a police officer and the body transporter. Both parties will then sign the property sheet.

If the clothing and property cannot be removed at the scene, it will be removed in the mortuary. Again two independent persons such as a pathologist and a police officer should witness the removal and sign the property sheets.

It should always be assumed the family would like the clothing and property returned, regardless of the state. On return, whether via the funeral director or to the family directly, details should be recorded on the property sheet and cosigned by the two parties.

Disaster Victim Identification

In the event of a mass disaster, body recovery can be crucial in assisting the final identification of the deceased person and reconstruction of the events just before the incident. The international disaster victim identification (DVI) forms published by Interpol are well recognized as the preferred method for documentation of body recovery and identification in the event of a disaster. The DVI protocol consists of five phases:

- Phase 1: The scene
- Phase 2: The mortuary
- Phase 3: Antemortem retrieval
- Phase 4: Reconciliation
- Phase 5: Debriefing.

Phase 1

Once the injured have been removed from the scene, the scene should be secured. DVI teams are formed. These consist of a crime scene examiner, a photographer and a recorder. A pathologist and a forensic odontologist support each team.

The location of each body or specimen is recorded such that it is related to a known reference point, usually a grid reference. A unique DVI number is given to each body or body part. The Interpol DVI form B is completed and the body or body part is moved to a body holding area. Property not attached to a body is also recorded with reference to its location and handed to a property officer.

During phase 1 the pathologist certifies death and assists with the identification of body parts. An odontologist may also be required

Phase 2

The body is transported to the mortuary in a body bag. Once at the mortuary, the body is radiographed and autopsied. The autopsy involves the assistance of DVI autopsy teams. These consist of a recorder, an examiner and a photographer.

With the assistance of the pathologists the DVI teams record and photograph all the clothing and property. Once the naked body has been photographed, they also record identifying features such as hair color, eye color, scars and tattoos. The autopsy is performed and the recorder again documents any items that may be utilized during the identification, such as the presence of an appendix. A forensic odontologist examines the teeth; trained fingerprint personnel take fingerprints. During the autopsy examination a sample of blood may be retained for DNA analysis These details are all recorded on DVI Interpol forms C1–G.

Phase 3

Trained police personnel interview family members of missing persons (presumed to have died in the mass disaster) to gather identifying features of that person. This includes the color and type of clothes that the

person may have worn, and other identifying features such as hair color, eye color, scars and tattoos. They are also responsible for gathering the dental records.

Phase 4

All antemortem forms and postmortem forms are compared during the reconciliation phase. This is achieved systematically using a reconciliation chart and grouping each set into male and female, black and white, and ages 0–15, 15–75, 75 years plus. If in doubt, the deceased is placed in the 15–75 group. The final identification of each case is presented to the coroner for the ultimate decision.

Phase 5

It is essential for all staff involved in the body recovery of a mass disaster to undergo debriefing. This could involve a hot debriefing immediately after the event and at the end of each working day, and could also involve a later debriefing, which also examines the procedures used and ways of improving these procedures.

See also: **Clinical Forensic Medicine**: Self-inflicted Injury. **Crime-scene Investigation and Examination**: Recovery of Human Remains; Preservation; Suspicious Deaths. **Fire Investigation**: Evidence Recovery at the Fire-scene. **Clinical Forensic Medicine**: Evaluation of Gunshot Wounds. **Anthropology**: Excavation and Retrieval of Forensic Remains.

Further Reading

Clark SC, Ernst MF, Hagland WD and Jentzen JM (1996) *The Mediocolegal Death Investigator*. Big Rapids, Michigan: Occupational Research and Assessment.
Geberth VJ (1996) *Practical Homicide Investigation*. Boca Raton, FL: CRC Press.
Knight B (1991) *Simpson's Forensic Medicine*. London: Edward Arnold.
Krogman WM and Iscan MY (1986) *The Human Skeleton in Forensic Medicine*, 2nd edn. Springfield, IL: Thomas.
Kendall RE (ed.) (1992) Disaster victim identification. *International Criminal Police Review* **October**:437–438.

PATTERN EVIDENCE

Contents

'Plastic' Bag Striations
Bare Footprint Marks
Footwear Marks
Serial Number
Shotgun Ammunition on a Target
Tools
Vechicle Tire Marks and Tire Track Measurement

'Plastic' Bag Striations

C Roux, S Bull, University of Technology, Sydney, Australia
S Olinder, Australian Federal Police, Melbourne, Australia

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0593

Introduction

Plastics in general, and plastic bags and films in particular, are in common use in a large variety of daily activities. As a result, they are frequently employed in relation to the commission of crime, the most obvious example being packaging material in drug trafficking cases. The ability to compare and ultimately identify

plastic bags or films from crime scenes with similar items in the possession of suspects, or found in serial cases, is of prime importance for forensic science providers. In general this examination relies on optical, physical and chemical techniques. The combination of identifying features allows plastic bags or sections of plastic film to be compared and conclusions to be drawn about their origin or relationship to each other.

It is the aim of this article to present and discuss the examination of plastic bags, and more specifically the use and value of some of their physical features, often known as 'plastic bag striations'.

Background

An awarness of the principles and processes underlying plastics chemistry and manufacture is necessary in order to appreciate the nature and value of the characteristics that will be encountered during an examination and comparison of such items.

Plastics chemistry

Plastic was originally developed as a cheap substitute for natural materials such as wood and metal. Since then, the industry has grown to produce plastics that are often superior to the products they were developed to replace.

Plastic is a byproduct of the petroleum industry, and the principal raw materials are organic chemicals used as monomers and plasticizers, and speciality chemicals that are used as additives to modify the properties of the plastic produced. Monomers are low molecular weight molecules which are joined together by strong covalent bonds, in a chemical process called polymerization, to form a chain of repeating monomer units. This chain has a high molecular weight and is called a polymer. Polymer resins are grouped by their monomer unit. The atoms along the center of the polymer chain are usually carbon atoms, and these form the backbone of the macromolecule. The atoms or groups of atoms on the side of the chain are called side groups or branches. One of the most simple and commonly used polymers is polyethylene, which is formed by the polymerization of the monomer, ethylene (**Fig. 1**).

Polyethylene, commonly called polythene, is versatile and can be either flexible or rigid, depending on its density. The density of a polymer is determined by the level of chain branching. As monomers link together, they sometimes form branches from the main chain. If there are only a few branches the polymer chains pack closely together, producing a high-density, stiff plastic. Chains with a greater number of branches lie further apart and produce a lower density plastic, which is therefore more flexible. Plastic bags and films are commonly made from a type of polyethylene. High-density polyethylene is used in the production of freezer bags, crinkly shopping bags and milk bottles, while low-density polyethylene is used to manufacture cling films and other articles, including garbage bins.

Like polyethylene, polypropylene is a simple polymer, formed by the polymerization of the monomer propylene (**Fig. 2**). It is also used in some clear films for food and microwave packaging.

Another more complex polymer is polyvinyl chloride (PVC), which is formed by the polymerization of the monomer, vinyl chloride (**Fig. 3**). Polyvinyl chloride is naturally rigid, but can be made flexible by the addition of a plasticizer, such as dioctyl adipate (DOA). Similarly, polyvinylidene chloride (PVDC), formed from vinylidene chloride, uses dihexylazelate plasticizer (**Fig. 4**). However, polymers which require the addition of plasticizers are being used less frequently in packaging materials traditionally used for food, as it has been found that some plasticizers may migrate into food products.

Polyethylene terephthalate (PET) is a relatively short-chain polymer with high strength and good resistance to acids, alkalis, bleaches, detergents and heat (**Fig. 5**). It is also a good barrier to carbon dioxide, and is therefore well known for its use in

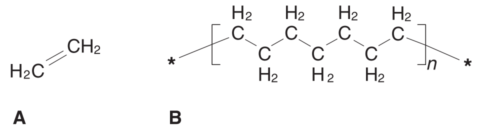


Figure 1 Structure of (**A**) ethylene and (**B**) polyethylene.

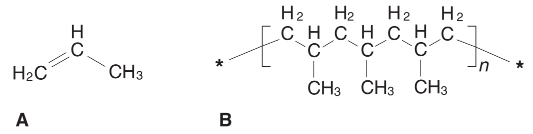


Figure 2 Structure of (**A**) propylene and (**B**) polypropylene.

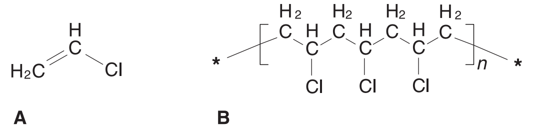


Figure 3 Structure of (**A**) vinyl chloride and (**B**) polyvinyl chloride.

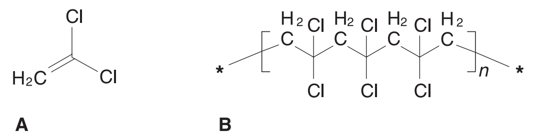


Figure 4 Structure of (**A**) vinylidene chloride and (**B**) polyvinylidene chloride.

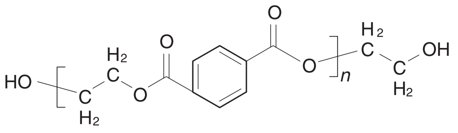


Figure 5 Structure of polyethylene terephthalate.

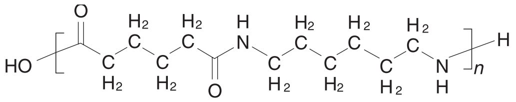


Figure 6 Structure of nylon 66.

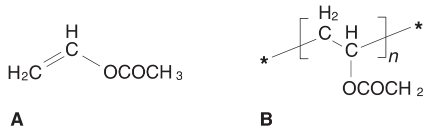


Figure 7 Structure of (**A**) vinyl acetate and (**B**) polyvinyl acetate.

carbonated soft-drink bottles. Another use is in cook-in oven bags.

The above polymers are all classified as thermoplastics. Thermoplastics are polymers which change form when heated, to allow molding into a desired shape. If thermoplastics are reheated, they return to a liquid form. This allows processes such as recycling to be carried out on these polymers.

Polyamide resins, or nylons, are polymers that consist of a repeating amide unit, and are known for their toughness and high temperature resistance (**Fig. 6**). Unlike the polymers mentioned thus far, once shaped they will not change shape or form, even when heated, and are therefore classified as thermosets. For this reason, they have also been used in oven bags.

A number of additives, such as fillers, stabilizers and pigments, are often mixed with polymers in a process known as compounding. These modifiers may increase the material's strength, reduce its cost, increase its heat and light resistance, impart flexibility, or add colour.

Copolymerization, which involves linking together two or more polymers, is sometimes carried out as an alternative method for changing the properties of the plastic. Ethylene vinyl acetate (EVA) is such a copolymer. EVA is a combination of polyethylene with polyvinyl acetate (**Fig. 7**). Polyvinyl acetate is resistant to oxidation and inert to the effects of ultraviolet and visible light, and when added to polyethylene it imparts these favorable characteristics. Because of these properties, EVA is used in plastic food packaging, where prevention of oxidation and light degradation are particularly important.

Plastic bag and film manufacture

The basic manufacturing process starts with virgin or reprocessed resin pellets. The resin pellets are placed in an extruder (see **Fig. 8**), which is cylindrically shaped, and then heated to approximately 250°C. Any pigments or other additives are then mixed in with the polymer to form a homogenous, viscous fluid. At this point, the polymer is forced through a screen filter attached to the end of the extruder. This serves to remove any large impurities from the mixture. After passing through the screen, there are two possible alternatives. The polymer can undergo either cast or blown extrusion.

Cast extrusion In cast extrusion, the polymer is forced through a straight die, in the form of a horizontal slit, which forms a flat sheet of film, with its thickness equal to the size of the slit (Fig. 8). It is then cooled, drawn off through rollers, and wound on to a roll.

Blown extrusion In blown extrusion, which is the more common type of extrusion, the molten plastic is extruded between a ring-shaped die and a mandrel, which is a rod at the axis of the die (**Fig. 9**). This type of extrusion forms a plastic tube. Pressurized air is blown through the mandrel to the inside of the plastic tube to allow the tube to be stretched to the desired size and thickness. The tube is then drawn upwards, cooled and flattened.

Generally, blown extrusion is used for plastic bag manufacture, but cling film may be manufactured using either blown or cast extrusion. If blown extrusion is used for cling film manufacture the extruded

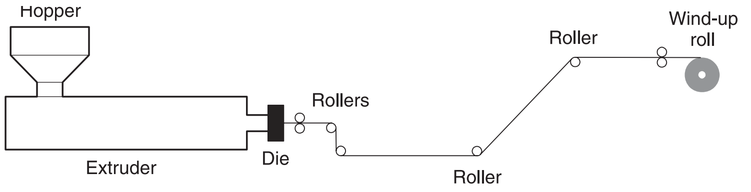


Figure 8 Cast extrusion.

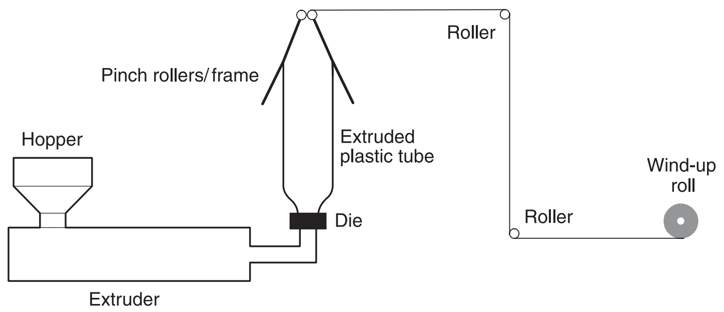


Figure 9 Blown extrusion.

tubular film is slit to form a single, wide sheet, which is then rolled on to a cardboard former and cut into individual rolls. To produce resealable bags by blown extrusion, the die has grooves in its sides to form the sealing strips.

After blown or cast extrusion, the plastic sheet is either wound on to a roll, or passed on for further manufacture, which may occur at a different location. The production of bags after extrusion varies greatly depending on the type of bag desired. The plastic sheet may undergo printing, folding, gusseting, addition of handles, heat sealing and cutting. The process becomes more complex as the bag itself becomes more complicated.

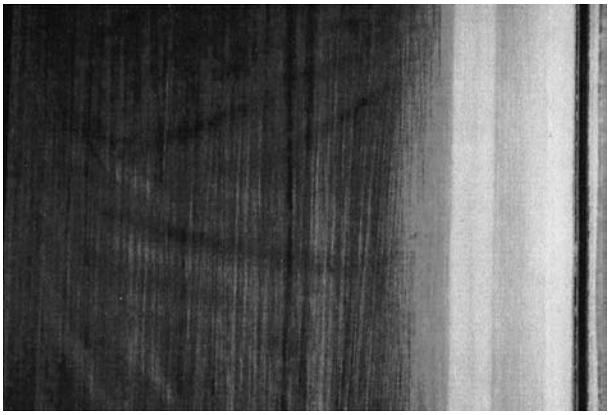


Figure 10 Striations ('die' and 'weld' lines).

Physical Features

A number of physical methods have been developed to examine plastics commonly used for packaging drugs. These physical methods generally depend on identifying characteristics that are permanently left on the bag or film during the manufacturing process. These identifying characteristics can be used to match plastic packaging to packaging found in a suspect's possession, or, more broadly, to show that the packaging comes from a certain batch or brand. Examples of such identifying characteristics are striations, pigment bands and 'tigerstripes'.

It is commonly observed that striations are left on the outer and inner surfaces of plastic film, from the die and mandrel, and rollers used to extrude the molten plastic. These are called hairline marks, and are separated into die lines (formed by damage to the die) and weld lines (formed by impurities caught in the die) (**Fig. 10**). If the die and mandrel are slowly rotated during production, the striations created will be slightly spiralling. On the other hand, if the die and mandrel are not rotated, the nonspiralling die lines may result in residues building up and wearing away, leaving slight random streaks or gauge variations in the plastic.

Pigment bands, which are caused by the inadequate mixing of dyes or pigments with the melted resin, may sometimes occur. They generally run in the direction of film production. These are useful characteristics for linking a large number of items to a common source.

'Tigerstripes', or horizontal streaks across the plastic film, may be formed during the stretching out of the film, by the pulling machinery, and the pressurized air used for this purpose (**Fig. 11**). These markings are useful for identifying items that are close to one another in the manufacturing sequence.

In addition, random impurities in the plastic, such as carbon material, resin, pigment or grit, that were

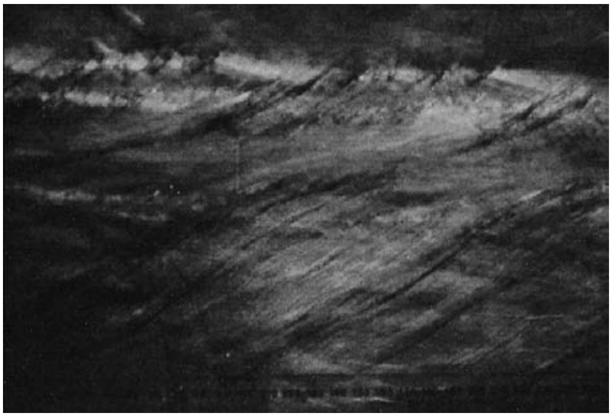


Figure 11 'Tigerstripes'.

not removed in the filtering step, result in dark or light spots of plastic, or 'fisheyes', in conjunction with radiating vertical streaks, known as 'arrowheads' (**Fig. 12**). Both characteristics are useful for identifying items manufactured close to each other in the manufacturing sequence.

Some plastics that have undergone a surface chemical treatment may have delicate surface characteristics, which are easily damaged features and are highly individual. Printing on plastic bags can sometimes be shown to have certain small defects, such as extra dots of ink, that can be used to show that the bags were manufactured on the same production line (**Fig. 13**).

Where heat sealers have been used to cut or seal plastic, it is possible to examine the distinctive impressions made in the seal by the fabric or Teflon covering the heated metal strips on the heat sealer. Also, cutting devices incorporated in sealing machines, or separate cutting devices, may leave characteristic snag marks or slight changes in direction of cutting which can be identified.

Where bags are torn from a roll by perforations, an examination of the perforations can be made. If a notched blade is used to make the perforation, it will consist of a line of cuts which looks like a dotted line, while if a saw-toothed blade is used the perforation will be a line of curved or V-shaped cuts.

Surface scratches and roller marks can be left on plastic bags either during manufacture, while being passed over rollers and other surfaces, or after manufacture, during handling of the bags. Scratches on two pieces of plastic can be compared side by side and can be used to identify a common source. Also, some cling films and plastic bags have distinct surface textures or patterns, which can be used to discriminate samples (**Fig. 14**).

Where pieces of plastic have been torn from a large piece of plastic film, or from a plastic bag, it may be possible to carry out a physical fit of the torn pieces.

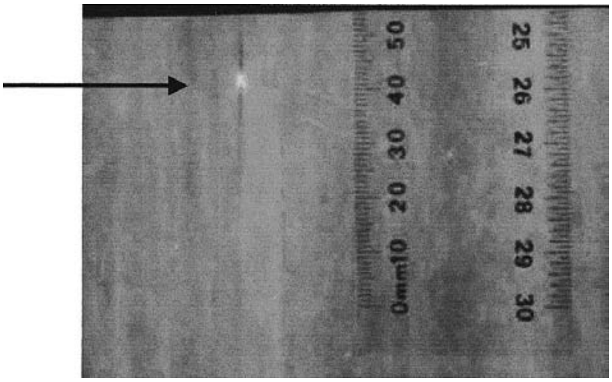


Figure 12 A 'fisheye'.

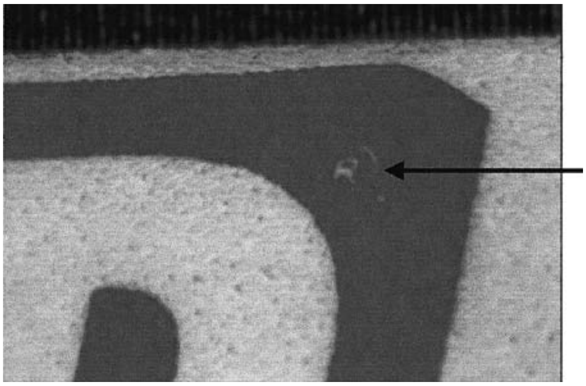


Figure 13 Printing defect on Glad Snap-lock Bag.

The combination of these physical features, along with the color, printing, dimensions, thickness, and the arrangement and type of sealing, cut edges and perforations (construction characteristics), allow plastic bags or sections of plastic film to be compared and conclusions to be drawn about their origin or relationship to each other.

Methods for Visualization of Physical Features

A number of methods have been described for visualizing the characteristic marks discussed above. These methods include:

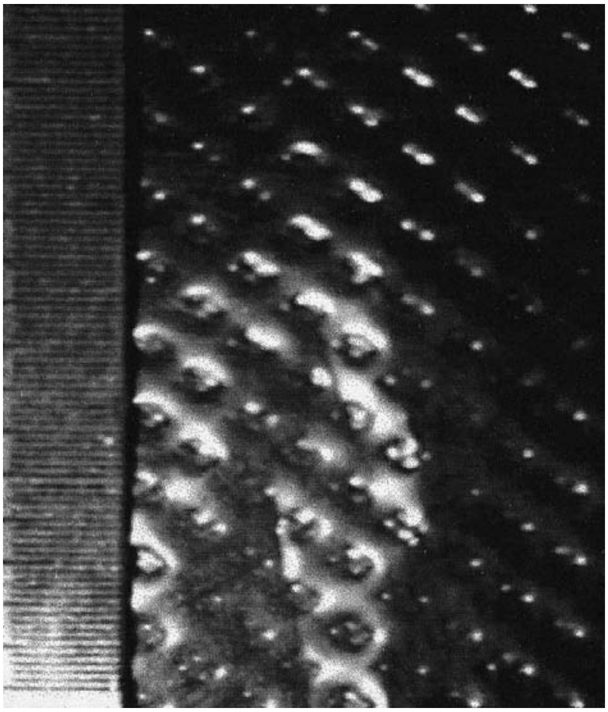


Figure 14 Surface pattern on sandwich bag.

- physical matching of cut or torn edges;
- transmitted, incident and oblique light;
- interference grid;
- shadowgraph photography;
- Schlieren photography;
- crossed polarizing filters;
- powdering.

Physical matching of cut or torn edges

There is no better evidence of common source than a demonstrated physical fit between torn edges of two samples of plastic. Matched cut edges can prove that two samples were manufactured consecutively in a sequence. This can strengthen the observations made using some of the other techniques.

Transmitted, incident and oblique light

These methods are very straighforward and constitute the early stage of any plastic bag examination. The only differences between the techniques are simply varaiations in the position of the light source. Transmitted light is useful to render visible characteristics, such as pigment bands and fisheyes found in pigmented plastic bags. Oblique transmitted light is best suited for transparent rather than pigmented films. Incident light reveals features on the surface of the plastic, such as scratches and printing. In all cases the characteristics are easily recorded by photography.

Incident and transmitted light microscopy can be carried out using a good quality stereomicroscope. In transmittance mode most of the light passing through the bag or film will not enter the objective lens, while light that is deflected by discontinuities, surface scratches or markings on a heat seal will enter the objective. With the use of incident microscopy it is possible to observe the details of any printing and the texture of the plastic.

Interference grid

In this method a grid of alternating black and white lines is illuminated behind a sample. The discontinuities in the plastic interfere with the appearance of the grid, allowing the characteristics to be visualized. These are then photographed and can then be compared. The advantage of this technique is that it is nondestructive and cost-effective, as the only equipment required is the black and white grid and a camera.

Shadowgraph photography

Shadowgraph photography is a simple method for comparing extrusion marks, heat-sealed edges, and cut edges on polythene sheet and polythene bags. The advantage of this method is that it is nondestructive, leaving the sample unchanged and available for further analysis.

A shadowgraph is made by light passing through the plastic film so as to make an image on a screen (**Fig. 15**). The inhomogeneities in the plastic film scatter or deflect the light out of its normal path to the screen, leaving a shadow.

Technically, the shadowgraph obtained on the screen can be recorded photographically or electronically. A high-contrast photographic film or paper can also be substituted for the screen. The use of a condenser-type photographic enlarger constitutes a variation of the technique. In this instance, a lighting system is used to focus on an area of the polythene sample. The area of interest is placed at the focal point of a secondary lens system which serves to project the focused image at a desired magnification. The image is then recorded on suitable photographic paper.

Schlieren photography

Schlieren photography is closely related to the shadowgraph technique described above. The difference in Schlieren photography is that the light scattered by the inhomogeneities is focused on the screen by the use of lenses or mirrors, together with knife-edge filters (**Fig. 16**).

Light from a point source passes through a collimating lens and is then collected and focused by a second lens using two spherical mirrors. The plastic film to be examined is placed in the area between the two lenses, which is known as the Schlieren field. A knife-edge filter is positioned at the focal point of the second lens and carefully adjusted across the beam of light, restricting the undeviated light while enhancing the light scattered by the inhomogeneities in the

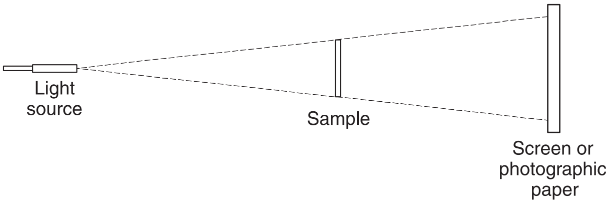


Figure 15 Shadowgraph photography.

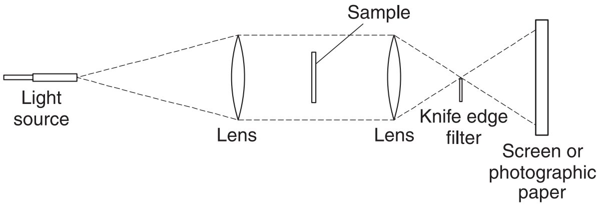


Figure 16 Schlieren photography.

plastic film. The image is recorded in the same way as for shadowgraphs.

Crossed polarizing filters

This technique takes advantage of the fact that polymer chains become oriented during the extrusion process. Two linear polarizing filters are placed at 90° to one another (crossed polarizing filters). The samples to be examined are placed between the two filters and rotated to an angle of 45° relative to the crossed polars. The samples show patterns of color if the plastic is birefringent, indicating variations in birefringence and/or in thickness. Machining marks present on plastic bags or film, such as scratches and roller marks, can also be visualized by this method. A large polarization table can be used so that multiple bags can be viewed. Alternatively, the patterns can be photographed and the photographs examined and compared at a later time.

Powdering

Some authors advocate the use of powdering to enhance surface features on plastic film. This can be achieved by carefully misting fingerprint black powder applied by blowing, or using a magenetic wand or a traditional fingerprint brush. However, this method can interfere with the search for fingerprints on the bags and a subsequent chemical analysis. For this reason, and also since other nondestructive methods exist, it appears sensible to avoid powdering, or at least apply the other methods first.

Chemical Methods

A wide variety of chemical methods can be applied to the examination of plastic packaging materials. These methods characterize the make-up of the plastic bags or films under investigation, including the polymer base, along with the dyes and pigments. They bring molecular and elemental information which is extremely valuable for comparison purposes. Common methods include ultraviolet-visible spectrometry, with or without microscope attachment, Fourier transform infrared spectroscopy, pyrolysis gas chromatography, thin-layer chromatography, X-ray fluorescence spectroscopy and neutron activation analysis. It is beyond the scope of this article to discuss these analytical techniques in detail; suffice it to say that chemical characterization constitutes a necessary step in the protocol for analyzing plastic bags or film. These methods can easily differentiate samples that appear visually similar in the first instance. However, it should be pointed out that chemical methods allow the samples to be identified at the class level only. The value of 'matching' chemical profiles is therefore assessed by the forensic scientist on the basis of the rarity and permanence of the major and minor materials making up the plastic bags, including the base polymer, additives, pigments and dyes.

Protocol for Forensic Examination of Plastic Bags and Films

The forensic examination of plastic bags and films follows the general stepwise method, from general to most detailed and from least to most destructive techniques. The actual individual methods applied greatly depend upon the resources of the laboratory, the case material (colored plastic, transparent plastic, minute sample, etc.) and to some extent the personal preferences of the examiner. However, in any case a stringent optical examination is applied first.

The first features to be taken into account are the general packaging type, the presence of any printing or labels, and the color and construction characteristics of the packaging. A description of any printing or labels should be recorded, along with the dimensions of the feature. The number of seams, folds, cuts, perforations and seals, and their position in relation to each other, should be noted. The dimensions of the packaging should be measured with a ruler. Appropriate measurements include length, width, depth (where appropriate), length of hem (portion of the bag past the bottom seal) and length above top seal. A general photograph of the packaging should be taken, including a scale. At this stage, close-up photographs may be taken of labels, printing and sealing mechanism.

A low-power stereomicroscope may be used to observe and photograph minute details, such as printing defects. The thickness of the drug packaging should be measured using a standard micrometer. The weight should be measured on an analytical balance, and a weight/area calculated.

A multiwavelength light source should be used to view colored or printed packaging, to help visualize any characteristic features, such as luminescence. The light source should be slowly scanned over the packaging using white light and light from 350 to 650 nm and any observations noted. This process is repeated numerous times, while viewing through a number of ultraviolet and colored filters. This step may be left out for colorless samples.

Some of the methods for visualizing physical features described above should then be used. To this end, crossed polarizing filters, which are relatively simple to use and which will usually provide extremely

high discrimination, are recommended. Both colored and colorless drug packaging should be viewed between two linear crossed polarizing filters, with transmitted light from a light box. Observations should be recorded and a photograph may be taken.

The use of shadowgraph and/or Schlieren photography is also recommended, given the value of these techniques for producing fine detail completely invisible to the unaided eye. These techniques are more complicated to set up, but are particularly useful where the plastic samples under examination exhibit poor birefringence colors under the crossed polarizing filters (a good example is cling film). By comparison the interference grid technique produces generally poorer results; however, it should not be ruled out as it has an obvious advantage in simplicity and low cost.

When the optical examination and physical measurements are completed, chemical analyses may be carried out. In this context, both molecular and elmental characterization are generally required. The choice of an individual technique mainly depends on the resources of the laboratory.

Value of Plastic Bag Evidence

There are a number of conclusions that may be drawn after completion of the sequence of analysis. These conclusions include whether the bags or pieces of film were or were not:

- manufactured from the same type of polymer;
- manufactured from the same chemical batch or formulation of a certain polymer;
- produced by the same machinery;
- directly connected during manufacture (i.e. adjacent on the same large piece of plastic film before cutting into bags or lengths of film).

A general caution should be heeded when carrying out any of the above analyses on plastic packaging. There are a number of possible ways that the product can be manufactured. The company that manufactures the endproduct may also produce the polymer resin from which it is manufactured, i.e. the company manufactures the product from start to finish. However, this is a rare situation. The company may buy the polymer resin and then manufacture the product. On the other hand, the company may buy lengths of plastic tubing, and then cut bags or lengths of film from this tubing. In some cases, the bags are bought premanufactured, and are simply packaged by the company before sale.

If all of these factors remained constant for a single product, there would not be a problem. The problem lies in the fact that the companies which produce plastic packaging often have several suppliers for the same ingredients or processes. Companies may change suppliers for a certain product or ingredient depending on cost, reliability and a number of other factors. Sometimes an entire process may be changed due to technological improvements in machinery, or chemical ingredients may be changed due to health, safety or environmental concerns (for example, the change from PVC to polyethylene in cling wraps).

This dynamic can be an advantage during the comparison process when the aim is to discriminate similar but unrelated samples. However, it is also a challenge during the early stages of the investigation when the identification of possible sources of manufacture is required and no comparison sample is available. This dynamic renders the setup of a database very difficult.

Nevertherless, plastic bag evidence can be extremely valuable when interpreted appropriately. For example, in a case where a large drug seizure has been made, and on searching a suspect's house or vehicle some plastic bags or film are found, it is then possible to compare the plastic packaging found containing the drugs to the plastic found in the suspect's possession. The visualization of machining marks can either show that the plastic packaging has completely different marks and is unrelated, or that the packaging has similar marks and has been manufactured on the same machinery at around the same time, or that the packaging was joined together during manufacturing, before being cut into bags or smaller lengths of film, and is therefore most probably from the same packet after manufacture.

However, it is important to be careful when making conclusions in this manner. All of the bags in one packet may not be in sequence. Also, during quality control procedures at the manufacturing plant, some bags may be removed from the packet for testing and replaced with bags from another packet or batch.

Therefore, if a plastic bag discovered in a suspect's possession is found to match a bag containing illicit drugs, it can be said that the bags were manufactured on the same machinery at around the same time, or, in some cases, that the bags were directly connected during manufacture, and were therefore from the same packet of bags. However, if the plastic bag discovered in a suspect's possession does not match the plastic bag containing illicit drugs, it cannot be said that the bags definitely did not come from the same packet.

See also: **Analytical Techniques**: Spectroscopy: Basic Principles; Presumptive Chemical Tests. **Photography and Digital Imaging**: Overview.

Further Reading

Castle DA, Gibbins B and Hamer PS (1994) Physical methods for examining and comparing transparent plastic bags and cling films. *Journal of the Forensic Science Society* **34**:61–68.

Chambers PA (1995) *Comparison of Plastic Wrap Materials Using Elemental Analysis Techniques*. Canberra: Australian Federal Police Research Project.

Cleverly B (1979) The comparison and matching of low density polythene plastic bags by infrared spectroscopy. *Journal of Forensic Sciences* **24**:339–345.

Denton S (1981) Extrusion marks in polythene film. *Journal of the Forensic Science Society* **21**:259–262.

Ford KN (1975) The physical comparison of polythene film. *Journal of the Forensic Science Society* **15**:107–113.

Gilbert J, Ingram JM, Scott MP and Underhill M (1991) The analysis of clingfilms by infrared spectroscopy and thermal desorption capillary gas chromatography. *Journal of the Forensic Science Society* **31**:337–347.

Griffin RME, Lewis R, Bennett J, Hamill J and Kee TG (1996) Analysis of cling films and coloured cellophanes and the application to casework. *Science and Justice* **36**:219–227.

Nir-El Y (1994) Forensic characteristics of coloured polyethylene bags. *Journal of Forensic Sciences* **39**:758–768.

Olinder SA (1995) *Characteristic Marks in Extruded Plastic Films*. Canberra: Australian Federal Police Research Project.

Pierce DS (1990) Identifiable markings on plastic. *Journal of Forensic Identification* **40**(2):51–59.

Roux C, Bull S, Goulding J and Lennerd C (2000) Tracing the source of illicit drugs through plastic packaging adelebose. *Journal of Forensic Sciences* **45**(1):99–114.

Smalldon KW and Moffat AC (1973) The calculation of discriminating power for a series of correlated attributes. *Journal of the Forensic Science Society* **13**:291–295.

Stanko RF and Attenberger DW (1992) The evidentiary value of plastic bags. *FBI Law Enforcement Bulletin* **61**:11–13.

Vanderkolk JR (1995) Identifying consecutively made garbage bags through manufactured characteristics. *Journal of Forensic Identification* **45**:38–50.

Bare Footprint Marks

R B Kennedy, Royal Canadian Mounted Police, Ottawa, Canada

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0540

Introduction

Many crimes are committed where the culprit must walk around the crime scene leaving telltale footwear impressions and at times, barefoot impressions, so there are a number of instances where the examination of a footprint recovered at a crime scene can be extremely important.

Forensic barefoot morphology involves the comparison of the weight-bearing areas of the bottom of a bare foot, when ridge detail is not present, in order to establish a link between the bare foot of an individual and an impression found in mud or some other medium, such as blood or the inside of a shoe, that may have been identified back to a crime scene. Often overlooked, but equally important, is the possibility of eliminating a suspect whose feet do not match the crime scene impression.

The most obvious example of the use of barefoot morphology is the comparison of a suspect's bare foot with that of an impression found at a crime scene. Also dependent on the uniqueness of a barefoot impression is the technique of comparing a foot to the impression found on the inside of a shoe. Crime scene footwear impressions exhibiting accidental characteristics can be positively identified to the shoe that made the impression. If a suspect is not found in possession of this footwear, it is still possible to link the suspect to the footwear and hence to the crime scene. This is accomplished by comparing the wear areas on the outsole of the shoe, the wear areas caused by the tops of the toes on the inside uppers of the shoe, and the darkened and indented sweat areas found on the insole of the shoe to the shoes the suspect may have been wearing at the time of arrest and with the inked barefoot impressions taken from the suspect.

History

Although limited research into the uniqueness of barefoot impressions has been carried out until recently, barefoot comparisons have been presented in court for many years. Barefoot evidence dates back to 1888, when a criminal by the name of LeDru was identified through the analysis of footprints.

In a case that went to trial in Canada in 1948, two brothers, Donald and William Kett, were accused in a series of breaking and entering offenses. After Donald was convicted, William claimed he was innocent and that the shoes that were matched back to the crime scenes belonged to his brother. However, the shoes were cut open and the marks inside compared to the feet of the two brothers. Based on this comparison, it was determined that William had worn the shoes and he was also convicted.

New Scotland Yard reported a case in 1953 in which a burglar left a pair of shoes behind at the scene of the crime. The prime suspect denied ownership of the shoes and willingly volunteered an old pair of his own boots for comparison. Both the boots and

the shoes exhibited the same unusual wear patterns on the outside. In order to compare the inside impressions, casting material was poured into the shoes and boots. The resultant casts were shown to be very similar and the person was convicted.

In 1962 in The Netherlands, a safecracker tried to discard the clothes he had worn while committing his crime by throwing them into a canal. The clothes, including a pair of shoes, were recovered. A prosecution expert compared the recovered shoes to shoes from the suspect and concluded they were worn by the same person. The defendant then hired his own specialist, but was dismayed when his witness agreed with the first. The defendant, who until this time had not agreed to have his feet photographed or printed, asked a third expert to examine his feet. Again to his dismay, this witness also agreed with the other two. The defendant was convicted.

In a case in New Jersey in 1981, a bloody socked footprint impression at the scene of a homicide was compared to the foot impressions of two suspects. One suspect had left a bloody fingerprint at the crime scene but was eliminated as the person leaving the bloody footprint. The second suspect's foot impression, when compared, was very similar to the crime scene print, leaving the expert to declare that there was a high probability that the second suspect also participated in the murder. Both were convicted.

While comparisons of barefoot morphology have been attempted by several people in recent years, problems have been encountered when these attempts have been made by people lacking proper forensic training and knowledge. The most notable example of this was the work of Dr Louise Robbins, an anthropologist from North Carolina. Dr Robbins authored one of the first books dealing with the subject of barefoot comparisons, but unfortunately much of her testing and casework has reportedly been surrounded by controversy.

Comparison of Barefoot Impressions

The comparison of barefoot impressions is treated as any other physical match and incorporates the same methodologies and scientific approaches. Clarity is one of the most important elements. All characteristics in a physical match must be analysed, compared and evaluated before a conclusion is reached.

Evidence

Evidence may be in the form of a shoe, or a bare or stockinged foot impression in blood, mud, dust or some other medium either found at or identified to a crime scene. The evidence must be photographed in position before removal. Evidence enhancement techniques may be required, such as blood enhancement chemicals, forensic lighting techniques or lifting techniques, for example gel lifts or electrostatic lifts, when the impression is found in dust. Strict evidence continuity (chain of possession) must be maintained for court purposes.

When exhibits are received they should be initialed and dated by the examiner to maintain the continuity chain of evidence. When making a footwear comparison, the outsole should be examined for common wear and, if any is present, it should be photographed. The top portion should be cut from the shoe and examined for any wear marks on the lining which may have been caused by the tops of the toes. Again, any marks found should be photographed. All foreign material (hair, fibers, soil, dirt, etc.) should be collected from the crime scene footwear and given to the investigator for possible analysis. The insole is then examined for marks that may be left by the weight-bearing areas of the foot. Any impressions found on the insole of the footwear should be photographed and may require enhancement. To date, the best enhancement method has been the use of specialized light sources such as laser, UV lamp, infrared lamp or forensic light sources. The light source that provides the optimum results should be used.

If the case involves a barefoot impression found on flooring at the crime scene, the impression may also be enhanced chemically or with one of the previously mentioned light sources, and photographed.

Photography

Impression evidence must be photographed using a scale in the photograph and the film plane must be parallel to the object being photographed. Color or monochrome film may be used, depending on the impression being photographed. Actual size (1:1) enlargements are prepared, together with traced overlays of both the known and the suspect impressions. Traced overlays are an invaluable investigative tool when comparing the location and shape of various characteristics or areas of the foot.

Known impressions

The collection of known impressions from individuals is the next step in the evaluation process. Upon arrest, the footwear worn by the suspect at the time of arrest should be seized and, if possible, other footwear owned by the individual should also be seized. The top and bottom of the suspect's feet should be photographed. This will assist the examiner in determining the placement of the toes and what, if any, damage had been done to the foot by the footwear. Molded

impressions should then be taken of the foot using conventional casting materials and methods.

Inked impressions should also be taken of the suspect's feet. It is important that both standing and walking impressions, with and without socks, are taken in order to see any discrepancies that might be present between the standing and walking impressions. When examining the walking impressions, all left-foot impressions should have similar shapes and size, as should all right-foot impressions. While all material, such as other footwear, molded impressions and photographs obtainable from a suspect, should be examined, it is preferable, but not necessary, to compare like impressions, i.e. comparing two-dimensional impressions with two-dimensional impressions and three-dimensional impressions with three-dimensional impressions.

When taking molded impressions of an individual's foot, a foam material, used by chiropodists (podiatrists), is the medium used, with the person putting weight on the foot and into the material. Dental stone is then poured in to fill the foam impressions, making a positive cast of the individual's foot (**Fig. 1**).

The inked impressions are taken from an individual on a piece of paper approximately 5–6 m long and 90 cm wide. An inking pad is made using an acetate-covered piece of cardboard large enough for a person to stand on. Fingerprint ink is spread over the acetate, using an ink roller to smooth out the ink. The individual stands on the inked pad, rocking back and forth to ensure the feet are inked properly. He or she then walks the length of the paper to the right side and walks back on the left side of the paper. Attention is paid to the individual's feet to ensure the person is walking normally.

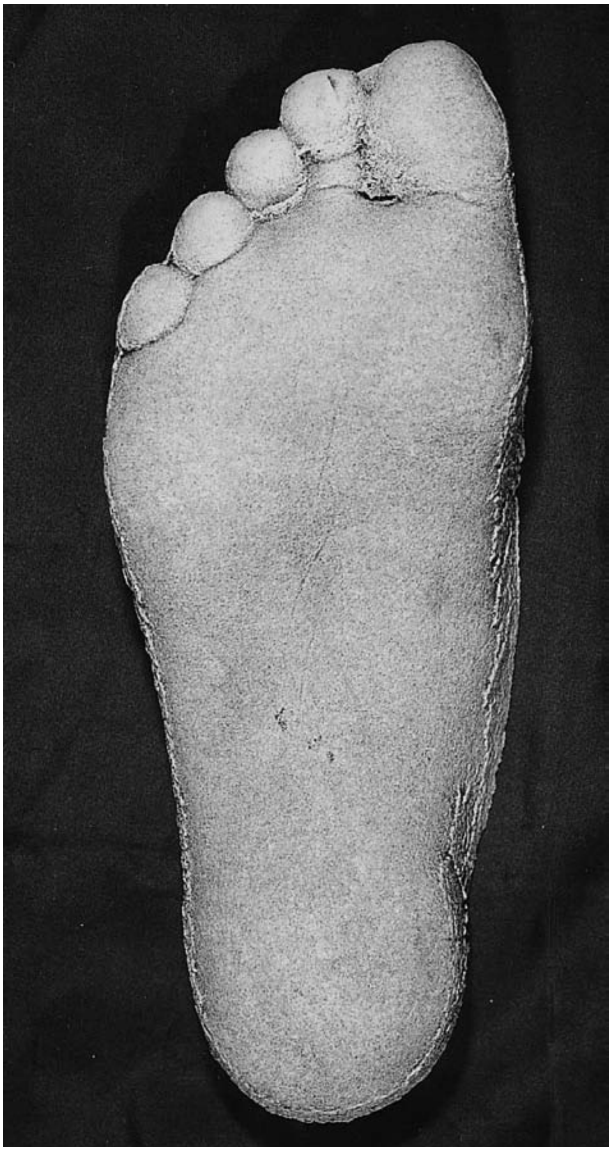


Figure 1 (see color plate 39) Dental stone cast, made from a foam impression material.

Barefoot comparisons

The first step in the comparison process is to examine what may be described as class characteristics. Class characteristics are those that are common to more than one foot; for example, the overall size of the foot, the fact that the correct number of toes are present, and the width of the ball and heel of the foot. The finer details of the foot impressions are then compared. For example, the examiner should look at the shape and placement of each toe, the relationship to all the other toes, the distance from the ridge that separates the ball of the foot from the toe pads, the contour of that ridge, the contour of the ball of the foot, and the shape and length of the arch area. While each feature individually may not be unique, the features in combination can make the impression unique. One must also evaluate any temporary changes in characteristics that may have been caused by injury to the foot, and which are not part of the permanent foot morphology. Accidental characteristics in sufficient numbers can be used to individualize the bare foot.

Barefoot impressions in shoes

We take thousands of steps every day, generating heat and sweat which cause the shoe to stretch and conform to the shape of the foot. The weight-bearing areas of the foot, including the toe pads, cause stains and indentations inside the shoe. It is these areas that are compared. If the footwear has damage, such as a nail coming through the insole, or the stitching causing a wrinkle on the inside of the footwear, then these abnormalities can injure the foot, causing accidental

characteristics. It is these characteristics that are used during the forensic analysis.

When the crime scene investigator is trying to establish a link between a barefoot impression of a suspect and an impression on the insole of a shoe which was found at or has been matched back to the crime scene, it is preferable to obtain a pair of shoes from the suspect for comparison. The comparison is started by examining the outsole of both the known footwear and the suspect footwear for similar wear patterns. The heel of each shoe is examined for wear. This wear should be consistent from shoe to shoe if the same person wore both shoes. The wear pattern on the ball of the foot is then compared and this also should be consistent in each shoe, if worn by the same person (**Fig. 2**). The shoe is then taken apart and the inside uppers and sides are examined for wear characteristics caused by either the tops of the toes or the sides of the toes and the ball of the foot, again looking for any damage that may have been caused by the foot (**Fig. 3**). The darkened and indented sweat impression found on the insole of the shoe is then examined and compared with the known shoe (**Fig. 4**). All the detail present in the shoe is examined in a similar manner to a barefoot comparison in order to form a conclusion.

Other comparisons

The identification of barefoot impressions can also be used at times of mass disasters where the identification of human remains may be impossible, e.g. when only a leg or foot may have been found. The foot or the

Figure 2 (see color plate 40) Outsole of a running shoe, showing wear areas on the ball of the foot and the heel.

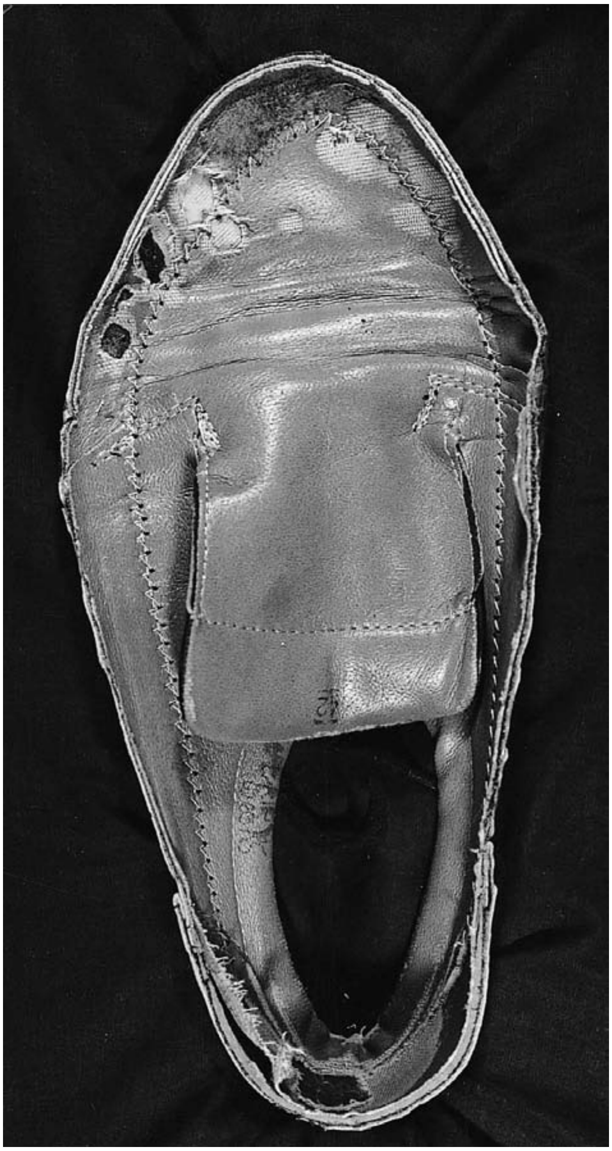


Figure 3 (see color plate 41) Inside top portion of a shoe, showing wear areas on the lining, caused by the top area of the toes.

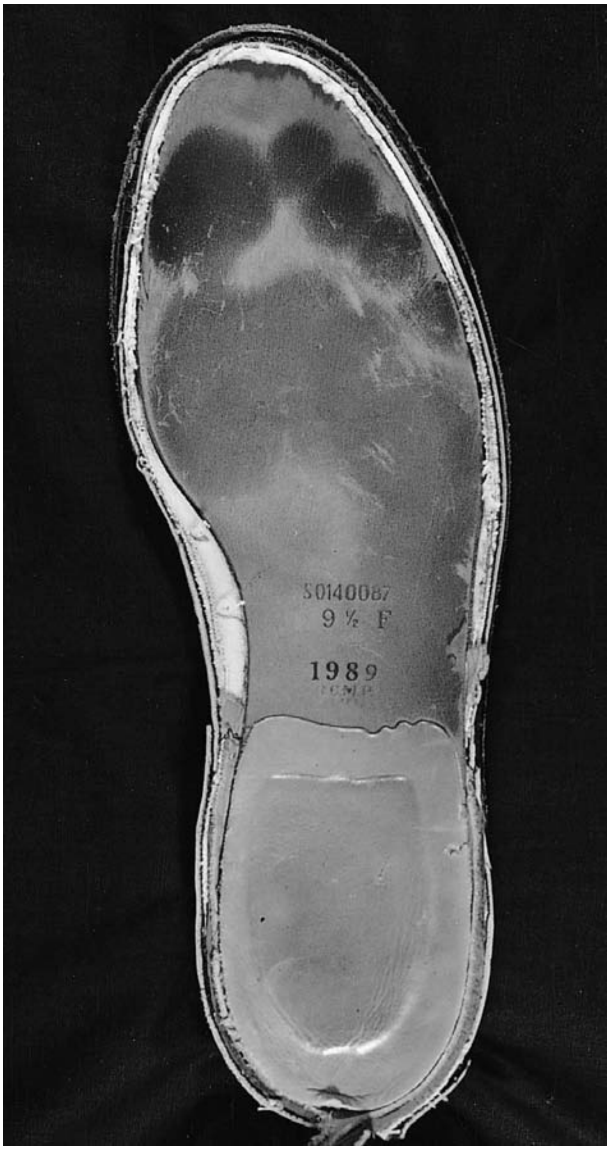


Figure 4 (see color plate 42) Insole of a shoe, showing the darkened and indented sweat areas caused by the weight-bearing areas of a bare foot.

footwear found on the remains can then be compared with a shoe taken from the victim's home in order to link the victim of this disaster to his or her residence. This procedure may also be used when trying to identify the remains of a missing person, when dental or medical records are not available because of the age of the victim, or other unknown reasons.

Research

As mentioned earlier, barefoot comparison evidence has been presented in court a number of times, even though research into the uniqueness of footprints is incomplete. It is hoped that current work using computerized databases and statistical analysis may finally answer some of the questions related to barefoot morphology and provide the ability to identify or eliminate suspects.

A great deal of early research and casework in barefoot morphology was carried out in India, probably due to the fact that people there would be more likely to walk around barefoot or in sandals. For example, Puri, in 1965, described his work of classifying and measuring barefoot impressions for comparison purposes. In 1980, Qamra, Sharma and Kaila published the results of their preliminary study involving the measurements of the footprints of 725 individuals.

In North America, Dr Louise Robbins carried out studies on the individuality of human footprints in the 1970s. Other forensic investigators have at times collected footprints, of the order of 100 samples, in relation to an ongoing case to establish to the judicial system an idea of the individuality of foot morphology. On the other hand, a large study of 6800 subjects, carried out in 1981–1982 by the Prescription Footwear Association, was undertaken with an eye to looking at how shoes fit, and the mismatching of feet for an individual, but not to show the uniqueness of barefoot impressions. Other collections of footprints have been taken in attempts to correlate foot measurements with height, weight, age or ethnic group.

The FBI has given evidence on barefoot morphology and has conducted considerable research in trying to establish the uniqueness of barefoot morphology. Their collection of impressions from 500 volunteers provided a starting point for studies of this nature. The Royal Canadian Mounted Police have been carrying out research on the uniqueness of barefoot impressions since 1989, and have extended this research considerably, with the collection of over 5000 pairs of barefoot impressions. This research attempts to prove the hypothesis that feet are unique to the individual by collecting inked walking barefoot impressions and entering the measurements of weight-bearing areas of the foot in a computer database. The computer program is capable of adding data to the system, and as entry of new data takes place, they are searched against all the data presently in the system to determine if a match exists. The system is capable of extracting data in any order for analysis by mathematicians and statisticians (**Fig. 5**).

Uniqueness of Bare Feet

Over time, as a person grows, the length and width of the foot, the shape and placement of the pads of the toes, the ball of the foot and the heel can change randomly and be affected by our environment. Before our bones ossify, they can be forced to grow

Measure in mm	L	R	Remarks			
Length	273.5	276.3				
Ball width	94.2	95.2				
Heel width	59.6	58.5	Border heel to			
Center heel to	center L	center R	border L	border R		
1st toe	226.4	226.7	273.5	276.3		
2nd toe	226.9	227.9	264.6	267.0		
3rd toe	215.8	218.6	253.8	257.3		
4th toe	202.1	203.0	238.6	240.9		
5th toe	186.4	189.9	223.1	227.3		
LMT	L			R		
	1	2	3	1	2	3
Border	223.8	221.5	203.5	228.7	226.9	205.7
Center	195.2	192.7	175.2	199.3	197.2	176.6
Name						
Race						
Height		Age				
Weight		Ethnic origin				
Shoe size		Sex				
Comment: none						
File Name: 00000005						

Figure 5 Printout from computer database showing the tracing of the weight-bearing areas of a pair of bare feet. The data box shows some of the data gathered by the computer and the history box shows the personal information of the individual. LMT, length of metatarsal.

improperly by the type of shoes worn, the type of activities we do or the injuries that sometimes happen over time. When growth stops, it has been shown that barefoot impressions will remain unchanged, unless affected by injury. In addition to injury to the foot itself, injury to the back, hip, knee or leg can cause the individual to walk differently and thus change the weight-bearing areas of the foot. Research has shown barefoot impressions over several years from the same adult person have remained unchanged. Impressions from inside several pairs of footwear that were worn over a 25 year period were examined and showed little change in the weight-bearing areas on the insole of the footwear. Many walking impressions taken from individuals over a distance of 5–6 m exhibit little or no change in the weight-bearing areas transferred to the paper.

Barefoot impressions have also been taken from several pairs of identical twins, and in each case their barefoot impressions were distinct one from the other (**Fig. 6**).

We are all genetically different and it is simply a matter of training ourselves to observe these differences in order to establish individualization. This fact is proven in the examination of our DNA, our fingerprints and even our facial features. Each of us is capable of observing features and forming an opinion of identification without realizing it. We start very early in life recognizing the features of our parents' faces. A typical human face has two eyes, a nose, a mouth and two ears, all in generally the same location, and yet everyone looks different, and we can recognize these differences because of many years of looking at faces. Our brain quickly analyzes, compares and evaluates each feature in order to reach a conclusion as to the identification of that person. This process is the same when doing any physical comparison. To the untrained eye, for example, a fingerprint is nothing more than a finger mark, but it is through the comparison process that a technician can see the uniqueness of the detail in the print and individualize a fingerprint to a particular person.

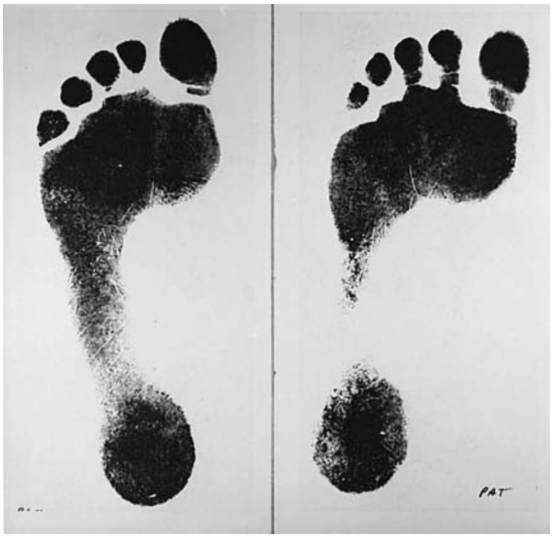


Figure 6 Inked barefoot impressions of the left feet of a pair of identical twins. This shows the differences between the placement of the toes, the presence of toe stems, and a high arch in one of the twins.

Initial research indicates that bare feet have characteristics that may form the basis for identification. While some research remains to be done to prove this hypothesis, barefoot impressions can still be compared in order to eliminate or link a suspect to the scene of a crime.

See also: **Basic Principles of Forensic Science**. **Crime-scene Investigation and Examination**: Collection and Chain of Evidence; Recovery of Human Remains; Packaging; Preservation; Contamination. **Evidence**: Classification; Statistical Interpretation of Evidence/Bayesian analysis. **Expert Witness**: Qualifications and Testimony. **Identification/Individualization**: Overview and Meaning of ID. **Pattern Evidence**: Footwear Marks.

Further Reading

Abbott JR (1964) *Footwear Evidence*, pp. 72–88. Springfield, IL: Thomas.

Bodziak WJ (1990) Wear characteristics. In: Geberth VJ (ed.) *Footwear Impression Evidence*, pp. 305–330. Amsterdam: Elsevier.

Cassidy MJ (1980) *Footwear Identification*, pp. 137–146. Ottawa: Royal Canadian Mounted Police.

Kennedy RB (1996) Uniqueness of bare feet and its use as a possible means of identification. *Forensic Science International Journal* **82**:81–87.

Robbins LM (1985) *Footprints*. Springfield, IL: Thomas.

Footwear Marks

W J Bodziak, Forensic Consultant Services, Jacksonville, FL, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0590

Introduction

Footwear leaves its marks and impressions through direct physical contact, in or on the surfaces over which it tracks. They are left in a large variety of two-dimensional and three-dimensional forms and are accurately sized representations of the footwear that caused them. Some impressions are more easily detected than others. Most are remarkably durable and contain sufficient detail for meaningful forensic examination.

Footwear impressions are most valuable as forensic evidence, in crimes of burglary, assault, robberies, rape, homicide and other crimes, where proof of an individual's presence is incriminating. They provide valuable information to the scene of crime officer and investigators, including information concerning the number of suspects, their path through the scene, their involvement in the crime, and their shoe type and size estimates. Impressions can reveal the point of entry and exit, often leading to other evidence.

The role of the forensic footwear examiner is to determine whether a particular item of footwear did or did not cause the marks or impressions recovered from a crime scene. This effort to either eliminate or identify the footwear evidence involves obtaining the best evidence from the scene of the crime, producing test prints from suspect footwear, conducting a laboratory examination and providing a written report. If necessary, the footwear examiner will offer this evidence and opinion in a court proceeding.

Forms of Marks

Two-dimensional footwear marks

The term 'footwear mark' is used to describe the resulting evidence an item of footwear leaves when it makes contact with a nongiving surface, such as a floor, kicked door or walkway. It is synonymous with the terms 'two-dimensional impression' and 'shoe print'.

The mechanics of a shoe making a two-dimensional impression through physical contact with a receiving surface is complicated by the fact that innumerable forms of trace residue are picked up, carried and later deposited by the shoe; these include all forms of dust, residue, grime, soil, blood, grease and oils, etc. This is further complicated because these varied materials combine with the features of virtually all porous and nonporous substrates, resulting in marks ranging from highly visible to latent, and from those that are highly detailed to those with minimal detail. The many possible combinations also complicate the process of choosing which is the best method for recovery.

Transfer of trace or residue materials With many two-dimensional impressions, there is a transfer of trace materials or residue between the footwear and the substrate. An impression that results when a shoe sole accumulates and then redeposits material is representative of the areas of the sole that actually came in contact with the ground surface. These are known as 'positive impressions' and account for most two-dimensional impressions. A 'negative impression' is produced when the contact areas of a shoe remove residue from a surface, leaving the residue where the

areas of the sole did not come in contact with the substrate. This occurs less frequently.

Two-dimensional impressions can occur when the shoes are wet or dry and can occur on a large variety of receiving surfaces. When a dry shoe leaves an impression on a dry surface, the impression is known as a 'dry origin' impression. The term 'wet origin' is used to describe an impression made when the shoe and/or surface contains moisture at the time the impression is made. Some examples of the many possibilities are the following.

Clean dry shoes on dry surfaces A shoe with a perfectly clean sole can still leave a trace of its impression on paper items, pieces of glass and polished surfaces by making an impression on that surface, or by removing the residue or film from those surfaces. For example, the physical contact of a clean shoe on paper items can result in impressions that can be developed chemically and physically.

Dry residue impressions When shoes track across a dirty surface, the bottom of the soles will accumulate a coating of residue. If they then track on to a relatively clean surface, that residue will be deposited in the form of footwear impressions.

Wet impressions Impressions made when a shoe sole is wet or damp constitute another category of impression. Most of these impressions dry out before they are found, but must still be considered as impressions of wet origin for the purposes of recovery. When there is rain, heavy dew or snow on the ground, most of the impressions, particularly at the point of entry, will be wet impressions. Many shoe soles retain moisture for a period of time and consequently leave wet origin impressions well inside interior areas. Impressions can also occur when the sole is wet and contains visible residue, such as mud or wet residue.

Depressed marks

The term 'depressed mark' is used to describe the evidence an item of footwear leaves when making contact with a deformable surface, such as soil, sand and snow. It is synonymous with the term 'three-dimensional impression'. Depending on the qualities of, and moisture present in, the substrate, the impression can be deep or shallow and it can retain detail ranging from excellent to one that cannot even be recognized as an impression of a shoe. Three-dimensional impressions marks are located primarily on exterior surfaces, but can include those found inside on deep-pile carpet, spilled materials and other deformable surfaces.

For simplicity, the more general term 'impression' will be used hereafter to refer to all forms of footwear evidence.

Information from Impressions

Footwear impressions provide valuable information. At the scene of the crime, they allow for the following.

Identification of footwear A specific item of footwear, to the exclusion of all others, can be identified as having left an impression at a crime scene.

Elimination of footwear Items of footwear can be eliminated as possibly having left an impression at the crime scene.

Linking footwear to a particular time Impressions left on objects moved during the crime, such as on paper removed from a safe, or on broken glass at the point of entry, on newly fallen snow, or in the victim's blood, provide restrictions on the time frame in which they were left and therefore relate the impressions to the crime.

Participation in the crime Footwear can be linked to a crime based on the time frame restrictions, and based on the location of those impressions and circumstances surrounding a particular event. They may also be linked to the crime if the owner of the footwear has no other explanation of how the footwear impressions otherwise legally occurred at the scene.

Location of impressions Impressions at the points of entry and exit provide a means of locating other impressions or evidence. Impressions also provide information about the path through the scene, and activities which occurred within.

Corroboration or rebuttal of alibis The presence and specific location of, or the absence of, impressions is used to corroborate or rebut the statements or alibis of suspected persons.

Classification Classification of the footwear impressions allows for the determination of the brand of manufacturer and style of the footwear that left the impressions. In most full-service laboratories this is done with a computer database, wherein the sole designs are arranged in a suitable manner for searching crime scene prints. This information often contributes to the development of a suspect.

Crime scene linking The same or a similar database is used by some laboratories to store impressions from crime scenes and successfully link crime scenes through the features of the crime scene impressions.

Estimate or determination of shoe size Crime scene impressions can also allow both estimates as well as precise determinations of shoe size. The linear dimensions of full or nearly full impressions allow for a general estimate of shoe sizing. A more specific method of sizing is possible if the manufacturer of the footwear is known, and comparisons can be made with samples of various sizes from that manufacturer. In this case, accurate sizing is possible with partial impressions as well as full impressions.

Estimate of height Based on the length of the footwear impressions, an approximate estimate of the perpetrator's height can be made. This estimate may have some investigative value.

Number of perpetrators involved Impressions of different designs or sizes of shoes at the scene can provide information about the number of persons involved in the crime.

Location and association with tire treads of perpetrator's vehicle Backtracking footwear impressions from the point of entry, or tracking impressions exiting the scene, can lead to the tire tracks of the perpetrator's vehicle.

Gait characteristics Gait analysis is concerned with the systematic study of human walking. Measurements of gait characteristics include stride length, step length, stride width and foot angle. Footwear impressions at crime scenes rarely present a succession of impressions that allows reliable use of this type of information. Further, it is concerned itself with general characteristics and is not suitable for personal identification.

Tracking Tracking involves the use of observing 'signs' to follow and locate an individual. This includes the use of shoe prints, bare footprints, crushed debris or displaced material that has been stepped on, and any other evidence of the passage of a person. It utilizes the gait characteristics to help in locating the various 'signs' found in successive steps. It is most commonly used for illegal aliens crossing borders, and missing children, but has also been used to locate criminals by following their tracks from the crime scene.

Treatment and Recovery

Footwear impressions at the scene of a crime are utilized in cases where proof of a person's presence is of some significance. The extent of how impressions can be used in an investigation and the results of the forensic examination depend critically on the treatment and recovery of those impressions. This process is easily divided into two steps: (1) locating and protecting the impressions; and (2) recovering the impressions.

Locating impressions

Most footwear impressions are on floor surfaces and are therefore subject to being tracked over by the shoes of other individuals. In order to avoid additional foot traffic over areas that may contain impressions of the perpetrator(s), the scene should be secured as quickly as possible. Details concerning the crime should be factored into the evidence recovery process. If footwear marks are of potential value in the proof of facts, attempts to aggressively search for, protect, preserve and recover those impressions should be set in motion.

Many impressions are obvious upon visual examinations of the scene. Others require special methods to assist in their location. These include darkening the room and searching for impressions with the use of high density oblique lighting and alternative light sources; the use of specular reflection to view the mark from the opposite side of a strong light source; and the use of electrostatic lifting in areas believed to have been walked on, such as the point of forced entry. Areas searched include not only flooring that may have been walked on but also any pieces of paper, broken glass or other objects on the floor, doors that may have been kicked in, and other surfaces, such as bank counter tops, desks, chairs, etc., that may have been stepped or walked on.

Exterior surfaces are searched with both existing and oblique lighting, and should be extended to include areas in the location around any forced points of entry, the escape routes and any remote areas where the perpetrator may have hidden.

All impressions should be considered to be of value. Partial and faintly visible impressions are likely to be as valuable as whole impressions. Areas or items containing impressions should be protected until the impressions can be properly recovered.

Recovering impressions

Once impressions are located, the preservation and recovery of those impressions is of paramount importance, as their detail will impact on any subsequent examination. Notes, photography, lifting, casting

and enhancement methods are all used to recover the maximum amount of information.

Written notes and documentation Notes providing information about each impression should be prepared. A description of the impressions (a residue impression on a piece of broken glass), their position and direction (at point of forced entry heading toward interior door) and substrate and/or weather conditions (exterior dew on grass, dry inside) only takes minutes to note. These observations are of value to the examiner in considering enhancement methods and may also have investigative importance later.

General scene photography General scene photographs are taken in series, from long to middle to close distances. Their zoom-in effect provides photographic documentation of the relationship of the impressions to the scene. General scene photographs are not intended for, nor do they contain the necessary detail for, use in forensic examinations of the footwear evidence.

Examination quality photographs Examination quality photographs are taken to recover the maximum amount of detail, specifically for use in a forensic footwear examination. To take examination quality photographs, a camera with a negative format of 35 mm or larger, and capable of manual focus, is placed on a tripod and positioned directly over the impression (**Fig. 1**). A finely divided scale, placed next to the impression, should be used in *every* exposure. The scale permits accurate photographic enlargement of the prints to a natural size, which is critical when conducting the examination. In the case of two-dimensional impressions, a very thin scale, at least 15 cm in length, should be set alongside the impression. In the case of three-dimensional impressions, the scale should be positioned alongside the impression and it must be set in the substrate on the same plane as the bottom of the impression. Oblique lighting should be used in the proper manner to increase the contrast in both the two-dimensional and three-dimensional impressions. Some three-dimensional impressions in light-colored substrates, such as snow and white sand, are particularly hard to photograph. For impressions in snow, Snow Print wax or an aerosol paint can be lightly sprayed at an oblique angle, from 50–60 cm away, to highlight the ridges or high spots of the impression. This will make the impression more visible, by adding contrast. This procedure is especially important with snow impressions because photographs of snow impressions alone provide little contrast and because subsequent casting of snow impressions is more difficult and less reliable. Very light application of colored aerosol paint of contrasting colors has also been used to highlight impressions in very light or very dark soils.

Figure 1 To take examination-quality photographs, a camera on a tripod should be positioned directly over the impression. A scale should be placed alongside the impression. A flash unit and long flash extension cord will allow the proper use of oblique light during exposure.

Original impressions Whenever possible, the original two-dimensional impressions, such as those on pieces of broken glass, paper, small rugs and the like, should be recovered and transmitted to the laboratory for examination. This should be done only after photography at the scene, and providing the impressions can be moved without damaging them. Moving them to the laboratory provides additional time and resources which often enable better recovery, enhancement and analysis. The removal of larger items containing impressions should be considered on an individual case basis.

Lifting techniques Lifting is performed (1) to improve the visibility and detail of the impression by transferring it to a surface providing better contrast; (2) to remove an impression from the scene to the laboratory; and (3) to search specific areas at the scene for latent footwear impressions.

Dry origin impressions To successfully lift a dry origin impression there must be certain differences between it and the surface it is on. A shoe tracking dust or dirt across a floor covered with the same dirt or dust does not leave a mark that can be separated by

lifting. If a shoe tracks from a dirty or dusty surface to a relatively cleaner floor surface, or steps on items such as paper, a chair seat or desk top, the impression could be lifted from that surface.

Wet origin impressions Wet origin impressions, as well as impressions in blood or greasy materials, are more firmly attached to and partially penetrate the surface. They are thus more difficult to lift. The use of cyanoacrylates, followed by powdering and lifting, or powdering alone followed by lifting, or the use of gelatin lifters on the original impression, are methods which provide some success.

Techniques or materials used to lift impressions include the following.

Electrostatic lifters These are devices that utilize a high voltage source to create static electricity, causing the transfer of a dry origin impression to a black lifting film. The black lifting film provides better contrast and allows for improved visibility of the impression. Photography of the lift with oblique light provides increased detail. Electrostatic lifting can be used to lift dry origin impressions from any surface (**Fig. 2**). In cases where attempts to lift the impression electrostatically are not successful, the process will not harm the impression, and will not affect the success of other, subsequent enhancement or recovery methods.

Gelatin lifters A gelatin lifter consists of a thick layer of gelatin and other components, laid upon a vinyl or cloth backing and covered with a protective clear cover sheet. Although less sticky than adhesive lifters, they can lift impressions from any surface, porous or nonporous. They are often used in lieu of the black electrostatic lifting process to lift dry origin impressions. They are also reasonably successful at lifting some wet origin impressions.

Adhesive lifters White adhesive lifters are used to lift powdered impressions. They can also be used to lift certain residue impressions from smooth surfaces, after which they are chemically treated to enhance the impression.

Microsil Microsil is a silicone product that can be used for lifting impressions. Its black or white forms are normally used for lifting powdered impressions on surfaces that are uneven or otherwise would not permit a complete lift when utilizing adhesive or gelatin materials. Microsil will release from most surfaces, dries completely, and can be lifted in 5 min.

Lifting odors from impressions Dogs trained to recognize and compare human scents are used in the traditional way for searching and tracking a scent line laid down by a human. In some countries, the scent is also collected from footwear impressions at the scene, with a specially prepared sterile cloth. The cloth is laid over the impression for a period of time, after which it is placed in a glass jar. A specially trained dog can then be used to associate this scent with the scents of several suspects in a line-up.

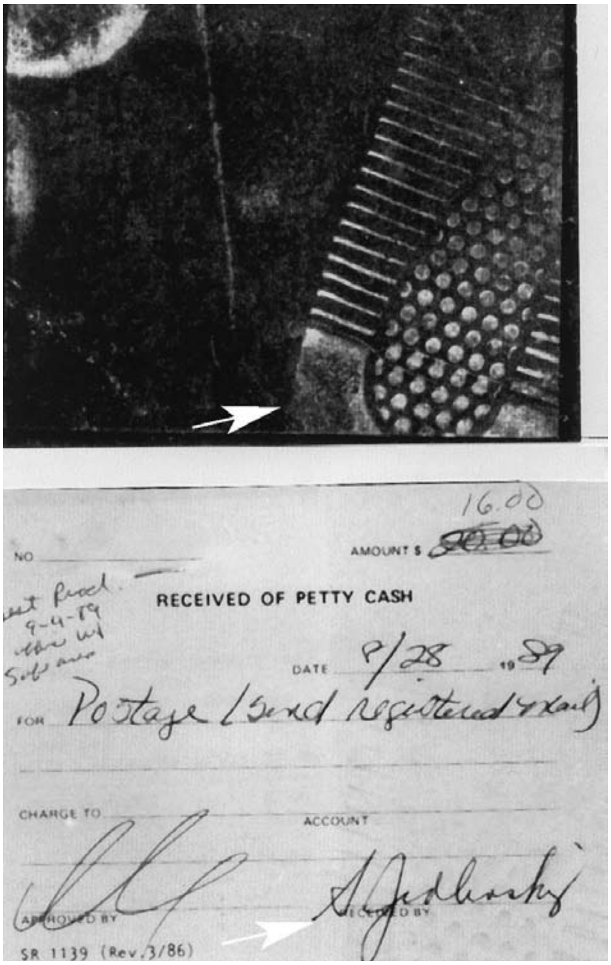


Figure 2 A residue impression is barely visible on a piece of paper knocked to the ground and stepped on during a robbery. Above it is a photograph of an electrostatic lift of the same piece of paper, clearly showing the footwear impression.

Casting techniques Casts back up photography and provide a three-dimensional representation of the impression. The cast permits evaluation of certain physical size and shape features that are normally not captured in a photograph. All three-dimensional impressions should be cast using dental stone.

Dental stone Dental stone is a gypsum-based product that has been modified for use in the dental industry. Dental stones, in contrast to the softer

plasters, have a high compressive strength and require a smaller water-to-powder ratio. Casting with dental stone provides a quick and easy way of recovering the three-dimensional features of an impression that has depth.

Casting with dental stone A popular method of using dental stone involves placing premeasured amounts, sufficient to fill a reasonably sized footwear impression, into reclosable plastic bags. When there is a need to cast an impression, a predetermined amount of water is added to the bag, which is then reclosed and used to mix the materials thoroughly. When completely mixed, the bag is opened, held at ground level, and poured so that the casting materials flows into the impression. The impression should be completely filled with the casting material until it overflows. The cast may be removed when hardened, usually in 20–30 min, or longer in colder conditions.

Release agents or fixatives *Release agents* are those materials, such as talcum powder or silicone spray, that are lightly applied to three-dimensional impressions to facilitate the release of the casting material from the substrate, resulting in a cleaner cast. They may be preferred by some but are not necessary with dental stone. *Fixatives* are those materials, such as dust hardeners or hair sprays, that are lightly sprayed indirectly on to fragile impressions to stabilize them, thus avoiding any degradation of the impression during casting. If not properly used, they may contribute more to harming the fragile impression then to preserving it. Most impressions do not require fixation.

Casting snow impressions Different snow types and variations in temperatures affect the amount of detail retained in impressions. Prior to casting, snow impressions should always be highlighted first with a light spray of Snow Print wax or colored spray paint. The spray should be directed at a low angle from a safe distance to avoid damaging the impression with the force of the spray. Once highlighted, the impression must be shaded from sunlight or bright indirect light as the colored material will accelerate melting. Examination quality photographs, with a scale placed next to the impression, but on the same plane as the bottom of the impression, should be taken. Snow impressions can be cast using Snow Print wax combined with cooled dental stone, cooled dental stone alone, or melted sulfur.

Enhancement of Impressions

Many impressions examined in the laboratory, or located at crime scenes, are indistinct, have poor contrast with the substrate or are altogether latent. A variety of enhancement techniques are routinely used to provide additional detail for comparison and, in some circumstances, as tools for developing and locating the impressions at the scene.

Specialized lighting and photography Specialized lighting and photographic techniques are nondestructive and are therefore used first in attempts to enhance impression evidence. Alternative light sources, high intensity lighting and oblique lighting are routinely used, at both the crime scene and in the laboratory, to provide increased detail and visibility of the impressions. The use of high contrast black and white film photography and the use of appropriate color filters with black and white film, for the purpose of selectively darkening the impression, are also routine. The additional use of computers, to increase the visualization of scanned original or photographed impressions, is also successful in increasing the contrast in some cases.

Physical methods These include (1) methods of transferring an impression to a lifting film that provides greater contrast; and (2) powdering with both conventional and fluorescent powders, on nonporous surfaces, to enhance or detect impressions.

Chemical methods There are many commonly used methods of chemically treating a footwear mark for detection or enhancement purposes. Those reagents that are used to enhance blood impressions include, but are not limited to, leuco crystal violet (**Fig. 3**), amido 10 black, diaminobenzidine, luminol, fuchsin acid and patent blue. Those reagents which are used to enhance residue impressions include, but are not limited to, physical developer, safranin 0, bromphenol blue, potassium thiocyanate, diazafluoren (**Fig. 4**) and 8-hydroxyquinoline. Cyanoacrylate, followed by powdering or dyes, is also used to enhance wet origin impressions.

Known Shoes of Suspects

Impressions recovered from the scene of the crime can include both those of the perpetrator(s) and those of the victim, police officers, medical personnel and other innocent persons. Once footwear impressions are recovered from a crime scene, comparisons to either eliminate or identify those impressions with shoes will follow. The actual shoes are needed to make known impressions for comparison, and to allow for a detailed and thorough examination of both class and individual characteristics.

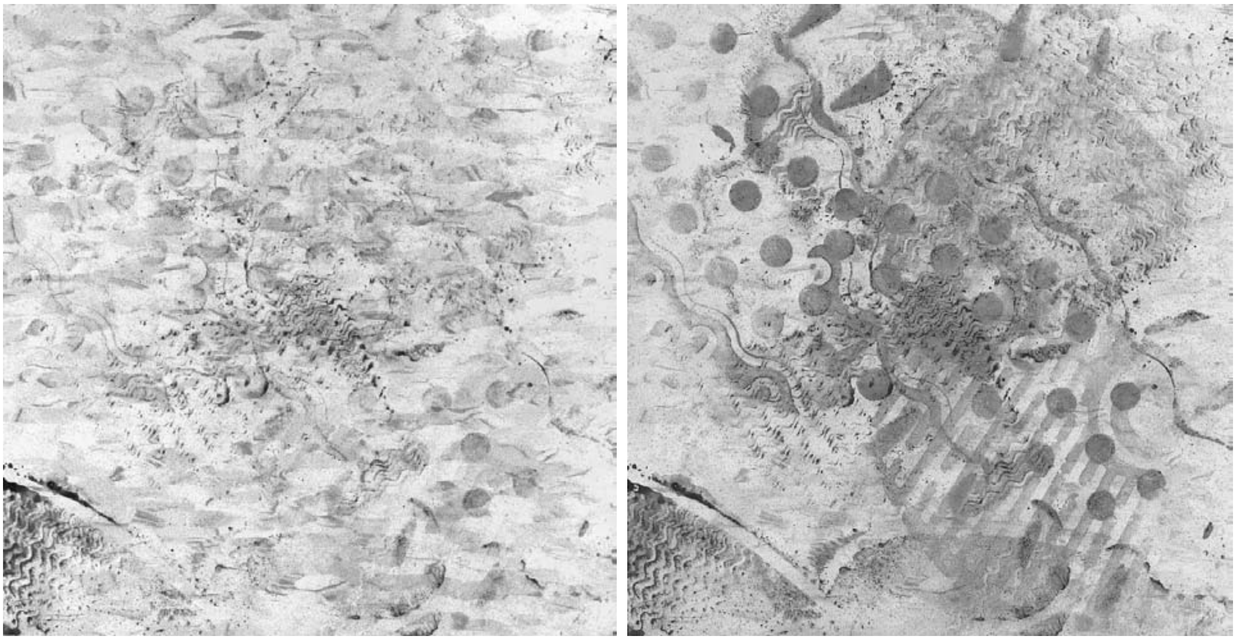


Figure 3 (see color plate 43) Leuco crystal violet is a clear, colorless solution, that can be easily applied at the crime scene or in the laboratory, for the enhancement of bloody footwear impressions. The first picture depicts multiple footwear impressions in blood on a light-colored tile floor. The second picture depicts the same impressions, but after enhancement with leuco crystal violet.

Footwear from suspected individuals

Footwear seized from suspected individuals, obtained within minutes from the time of the crime, might appropriately only include the shoes that the individual is wearing. However, persons normally own more than one pair of shoes, and may even possess several pairs of shoes that are very similar in design. Additionally, impressions of additional designs, not known to the investigator at the time of the seizure of the shoes, may be detected or developed on evidence in the laboratory at a later date. To avoid any chance that the wrong footwear is seized, footwear obtained from suspects, hours, days or weeks later, should include not only the footwear the suspect is wearing, but *all* footwear in that individual's home or vehicle.

Eliminating footwear of nonsuspects

Elimination examinations are important, in some instances, to insure that unaccounted footwear impressions are not those of the perpetrator or accomplice. Although it may not be possible to account for all impressions recovered, consideration as to whether a case warrants obtaining either a photographic record or elimination prints of shoes of all persons at the scene should be made early on. Photographs or test prints of footwear from known innocent persons at the scene are satisfactory for elimination purposes, based on differences in design and other class characteristics with the questioned impressions. Only if the elimination footwear is coincidently similar to the design of a questioned impression is it necessary to have the actual shoes.

Known impressions

A variety of methods exist for creating both two- and three-dimensional known impressions of suspected footwear for use in the comparison with the scene impressions. They include methods that utilize inks, fingerprint powders, casts and other materials. Known impressions should reflect a high degree of detail and do not need to attempt to duplicate the exact conditions of the crime scene impression. The known impressions are often made in, or converted to, a transparent form, so that they can be used as a superimposition tool by the examiner to assist in the evaluation of class and individual characteristics.

Abandoned or discarded footwear

Most footwear is removed from the feet of a suspect when he or she is apprehended or interviewed. Other shoes, which a suspect admits to owning, are found in the suspect's house or car. However, occasionally footwear used during a crime is abandoned or intentionally discarded by the perpetrator, to be located later during the investigation. If that footwear can be linked to the crime scene through its impressions, or the presence of the victim's blood, then it becomes important to associate or disassociate those shoes

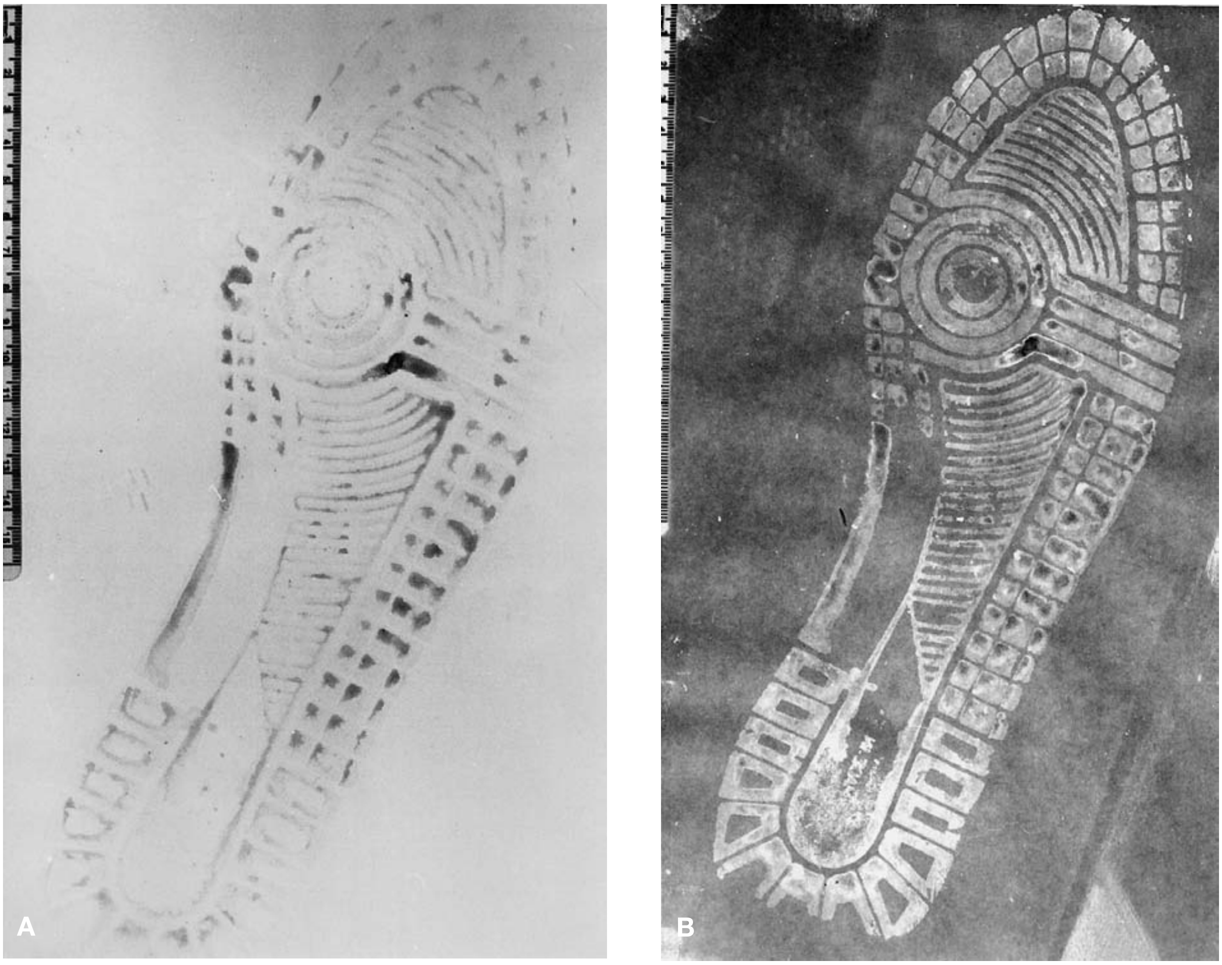


Figure 4 (A) A muddy print on paper before treatment. **(B)** The same print enhanced with DFO and photographed with an alternative light source. The print was lifted, and activated, with a black gelatin lifter. Courtesy of M J M Velders, Eindhoven, The Netherlands.

with the feet of potential suspects. Comparisons can be made between that footwear and a suspected owner's feet.

Method of Comparison

Basis for comparison

The physical contact of an item of footwear with a substrate results in the direct physical transfer of class and individual characteristics of that footwear to the substrate. Those characteristics, when retained by the substrate, accurately represent the class and individual characteristics of the footwear that made them. Those characteristics can be compared with footwear alleged to have made them.

Examination

Examination involves the comparison of the questioned crime scene impressions with the known shoes of suspects by a competent and qualified forensic footwear examiner. The forms of the questioned impressions examined can include original impression evidence, casts or lifts of the scene impressions, and enlarged natural-size photographic prints of the examination quality photographs of the impressions. The comparison utilizes side-by-side and superimposition methods, assisted by low magnification, specialized lighting and known impression exemplars of the shoes in question. Similarities or dissimilarities between the class, wear and individual characteristics form the basis for the examiner's resulting opinion. Results are dependent on the extent of individuality of the shoe as well as the quality of reproduction of detail in the crime scene mark or impression.

Comparison of both partial and full impressions is made with the respective areas of the suspect shoe. The areas examined include:

Design The design or pattern on the sole bottom is the first area of comparison. There are many thousands of shoe designs and those designs are constantly

being added to, while others are discontinued. The specific details of each design may also vary from mold to mold, due to the specific manufacturing process, thus adding to the already large number of possible choices. In addition, many designs also incorporate texture that is applied through a hand stippling or acid etch process. The texture permits even further distinction between molds. If the examination reveals the specific design is different from the crime scene impression, the suspect shoe could not have possibly made the impression and an elimination can be made. If the specific design corresponds with the respective area of the suspect shoe, the shoe cannot be eliminated and the examination proceeds. A conclusion that the design corresponds is very significant, because of the large number of choices of shoe designs.

Physical size and shape Each shoe design is made in many sizes. The soles of different sizes will have different dimensions and/or proportions throughout the size range. Further, different styles of shoes are made with different-shaped soles, allowing for further distinction, even between shoe soles of the same general length. The subsequent physical differences in the size and shape features of the sole is of further comparative value in associating or disassociating scene impressions with suspect footwear. It is noted that this portion of the footwear comparison pertains to the actual physical size and shape, and whether these correspond with the scene impression. It is not about the manufacturer's shoe size designation. In cases where a plain sole bottom does correspond in its physical size and shape with the scene impression, this constitutes a limited but still meaningful conclusion. In those cases where a shoe sole has a design, and both its design and physical size and shape correspond with the scene impression, the association is highly significant. This is because of the extremely large number of possible design and size and shape combinations, and the obvious fact that any specific shoe design/size combination would represent only an extremely small fraction of 1% of the overall shoe population. It should be noted here that minor discrepancies in the physical dimensions of recovered impressions, when compared with known impressions of suspect shoes, are normal. These occur for a variety of reasons resulting from the normal impression-making process and the recovery process. Minor discrepancies are insignificant. Larger discrepancies are cause for concern.

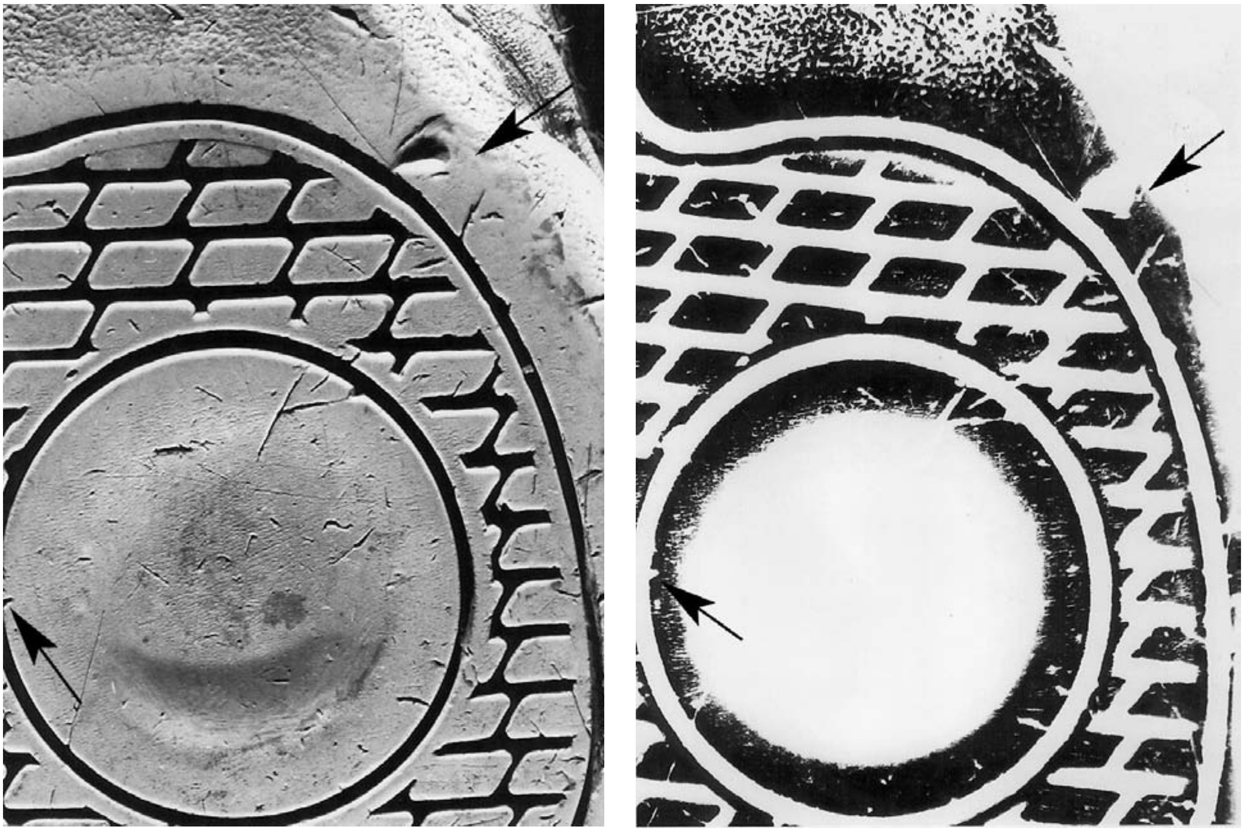
Wear As shoes are worn, their soles are altered by the abrasive forces created each time those soles make contact with the substrate. These abrasive forces slowly remove selected areas of the rubber. The areas that contact the surface with every step, and where pressure and movement of the foot are transmitted, receive the most wear, while areas that do not contact the surface receive little or no wear. As wear progresses, and the abrasive action removes the rubber, it alters portions of the design. The difference between the worn and unworn areas becomes obvious and is sometimes referred to as the wear pattern. As long as the shoe continues to be worn, the degree of wear and the wear pattern will gradually change. Wear is a function of time and, therefore, as a routine matter, the date of crime scene impression and the date suspect shoes are seized should be known by the examiner. The degree of correspondence in wear between a crime scene impression and the perpetrator's shoe depends on the amount of additional wear the shoe has received.

Wear is examined to assess: (1) the position of the wear, i.e. does the position of wear on the shoe match that of the crime scene impression; and (2) whether the degree of wear corresponds, which is more relevant in cases where the shoes are obtained shortly after the crime, versus shoes that may be obtained months later. The correspondence of wear characteristics offers an additional means of eliminating other possible shoes of the same size and design that are owned by members of the general population but which exhibit different wear. Therefore, correspondence of wear further reduces the possible number of shoes that could have made the impression. In some cases, when shoes and impressions exhibit very little or no wear, when the detail or other limitations in the recovered impression preclude an assessment of wear, or when a significant period of time has elapsed between the date of the crime scene impression and acquisition of the suspect's shoes, the relevance of wear in an examination may be minimal or insignificant. The absence of the ability to utilize wear during the examination in no way minimizes the conclusions that may result from comparison of the other aspects, i.e. size, design and identifying characteristics. The general wear features of footwear, even if matched precisely with the scene impression, are not a basis for identification of the shoe with the impression. When present, however, wear provides a significant contribution to the overall basis and results of the examination. In some cases, wear can be used to eliminate a shoe as having made an impression, should its condition of wear be clearly different, and should it be impossible for it to have been in the same condition of wear as that producing the crime scene impression. The wearing down of the sole can also result in the erosion or creation of individual identifying characteristics. For instance, as a shoe continues to wear, any

individual identifying cuts, scratches, etc., can be altered or altogether lost. At the same time, if a shoe wears excessively, the wear may create individual identifying characteristics, such as the rough edges of a hole worn through the sole, or may produce microscopic abrasion wear patterns which are highly unique.

Individual identifying characteristics Individual identifying characteristics include virtually anything that has been added to or removed from the shoe sole in a random fashion, thereby contributing to or giving a shoe sole uniqueness, and allowing it to be distinguished from other shoes soles. They most commonly include cuts, scratches, gouges, etc. sustained by the surface of the sole as the shoe is worn (**Fig. 5**). They also include materials that randomly become attached to the shoe sole, such as tar, gum, tacks or nails, stones imbedded in the design, etc. What makes individual identifying characteristics valuable is the fact that they occur randomly and are therefore unlikely to be in the same position on another shoe of the same design. It would be a remarkable coincidence to find even one similar feature in the same position on another shoe of the same design. Even more significant are the features of the characteristics, e.g. a V-shaped scratch, an irregularly shaped hole, even further individualized by its orientation (direction) and size. These features add incredible individuality to an item of footwear and serve to distinguish that item of footwear from all others. Individual identifying characteristics can also be transferred to shoe soles during the manufacturing process, but these are less frequent, and their value is often more difficult to assess, as in some instances this might require assistance from a particular manufacturer.

When a crime scene impression and a suspect shoe share sufficient random individual identifying characteristics, a positive identification is possible. A positive identification means that only that shoe, and no other shoe in the world, made the crime scene impression. There is no minimum number of random individual characteristics needed to establish an identification: one characteristic alone could be used to



A **B**

Figure 5 An enlarged area of **(A)** a shoe and **(B)** the respective known inked impression of that shoe, depict the many individual cuts and scratches. Individual random cuts, or groups of cuts, for example adjacent to the arrow on the top right, reflect sufficient size, shape, orientation and position features to make this shoe sole unique. Others, like the cut by the arrow on the left, although highly significant and important, reflect fewer features and would not be used alone to identify a shoe.

identify a shoe, providing the characteristic was sufficiently clear and detailed, and reflected sufficient significant features in common with the scene impression. An example would be a feature generally shaped like an 'X' with one stem of the 'X' longer than the other, and terminating in a slight curve. If a characteristic such as this were of the same size, shape, position and orientation on both a crime scene impression and a shoe sole of the same design, it could be sufficient alone to permit a positive identification. However, should a characteristic be less clear or contain fewer features, such as a very tiny hole, more than one characteristic would be needed for identification.

The Footwear Impression Expert

Footwear impression examinations should be made by persons who have received additional specialized training and are experienced in that discipline. Being qualified as an expert in other forensic disciplines is not, in itself, qualification to conduct examinations between questioned footwear impressions and known shoes. Examiners in this field should be afforded specific training and experience under the direct supervision of an experienced and qualified footwear impression examiner. This normally includes working cases and practical exercises of sufficient quantity over an extended period of time to enable the trainee examiner to acquire the necessary skills and knowledge related to this forensic discipline. Reading the available literature, attending specific courses, seminars and workshops, touring footwear manufacturing facilities and conducting research are ways of obtaining further knowledge.

See also: **Photography and Digital Imaging**: Overview.

Further Reading

Abbott JR (1964) In: German AC (ed.) *Footwear Evidence*. Springfield, IL: Thomas.

Bodziak WJ (1999) *Footwear Impression Evidence*, 2nd edn. Boca Raton, FL: CRC Press.

Cassidy MJ (1980) *Footwear Identification*. Ottawa: Canadian Government Publishing Centre. (Reprinted by Lightning Powder Company, Salem, OR.)

Giles E and Vallandigham PH (1991) Height estimation from foot and shoeprint length. *Journal of Forensic Sciences* **36**:1134–1151.

Girod A, Champod C and Margot P (1998) *L'Exploitation des Traces de Chaussures en Criminalistique*. Lausanne: Institut de Police Scientifique et de Criminologie, Université de Lausanne Série Criminalistique XI.

Mikkonen S, Suominen V and Heinonen P (1996) Use of footwear impressions in crime scene investigations assisted by computerized footwear collection system. *Forensic Science International* **82**:67–79.

Paine N (1998) Use of cyanoacrylate fuming and related enhancement techniques to develop shoe impressions on various surfaces. *Journal of Forensic Identification* **48**: 585–601.

Sharma BR (1980) *Footprints, Tracks and Trails in Criminal Investigation and Trials*, pp. 61–77. Allahabad: Central Law Agency.

Shor Y, Vinokurov A and Glattstein B (1998) The use of an adhesive lifter and pH indicator for the removal and enhancement of shoeprints in dust. *Journal of Forensic Science* **43**:182–184.

Theeuwen ABE and Limborgh JCM (1998) Enhancement of footwear impressions in blood: comparison of chemical methods for the visualization and enhancement of footwear impressions in blood. *Forensic Science International* **95**:133–151.

Whittle MW (1996) *Gait Analysis*, 2nd edn. Oxford: Butterworth-Heinemann.

Serial Number

C Petterd, Australian Federal Police, Canberra, Australia

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0595

Introduction

Serial numbers are present on many everyday objects. They are used to distinguish items from others of a similar appearance. In most instances, they are simply a sequential number, which describes how many units in a particular production sequence have been produced. In some cases, they are encoded to carry information about the product, such as date of manufacture, model type, etc.

Serial numbers are regularly removed from items in order to hide their true identity. The reasons for the removal are varied, and include:

- the perpetration of fraud, such as portraying an item to be of a different value when making an insurance claim;
- to hide the origin of items used in the commission of a crime, such as removing the number from a firearm used in a murder;
- to prevent positive identification of stolen property.

identify a shoe, providing the characteristic was sufficiently clear and detailed, and reflected sufficient significant features in common with the scene impression. An example would be a feature generally shaped like an 'X' with one stem of the 'X' longer than the other, and terminating in a slight curve. If a characteristic such as this were of the same size, shape, position and orientation on both a crime scene impression and a shoe sole of the same design, it could be sufficient alone to permit a positive identification. However, should a characteristic be less clear or contain fewer features, such as a very tiny hole, more than one characteristic would be needed for identification.

The Footwear Impression Expert

Footwear impression examinations should be made by persons who have received additional specialized training and are experienced in that discipline. Being qualified as an expert in other forensic disciplines is not, in itself, qualification to conduct examinations between questioned footwear impressions and known shoes. Examiners in this field should be afforded specific training and experience under the direct supervision of an experienced and qualified footwear impression examiner. This normally includes working cases and practical exercises of sufficient quantity over an extended period of time to enable the trainee examiner to acquire the necessary skills and knowledge related to this forensic discipline. Reading the available literature, attending specific courses, seminars and workshops, touring footwear manufacturing facilities and conducting research are ways of obtaining further knowledge.

See also: **Photography and Digital Imaging**: Overview.

Further Reading

Abbott JR (1964) In: German AC (ed.) *Footwear Evidence*. Springfield, IL: Thomas.

Bodziak WJ (1999) *Footwear Impression Evidence*, 2nd edn. Boca Raton, FL: CRC Press.

Cassidy MJ (1980) *Footwear Identification*. Ottawa: Canadian Government Publishing Centre. (Reprinted by Lightning Powder Company, Salem, OR.)

Giles E and Vallandigham PH (1991) Height estimation from foot and shoeprint length. *Journal of Forensic Sciences* **36**:1134–1151.

Girod A, Champod C and Margot P (1998) *L'Exploitation des Traces de Chaussures en Criminalistique*. Lausanne: Institut de Police Scientifique et de Criminologie, Université de Lausanne Série Criminalistique XI.

Mikkonen S, Suominen V and Heinonen P (1996) Use of footwear impressions in crime scene investigations assisted by computerized footwear collection system. *Forensic Science International* **82**:67–79.

Paine N (1998) Use of cyanoacrylate fuming and related enhancement techniques to develop shoe impressions on various surfaces. *Journal of Forensic Identification* **48**: 585–601.

Sharma BR (1980) *Footprints, Tracks and Trails in Criminal Investigation and Trials*, pp. 61–77. Allahabad: Central Law Agency.

Shor Y, Vinokurov A and Glattstein B (1998) The use of an adhesive lifter and pH indicator for the removal and enhancement of shoeprints in dust. *Journal of Forensic Science* **43**:182–184.

Theeuwen ABE and Limborgh JCM (1998) Enhancement of footwear impressions in blood: comparison of chemical methods for the visualization and enhancement of footwear impressions in blood. *Forensic Science International* **95**:133–151.

Whittle MW (1996) *Gait Analysis*, 2nd edn. Oxford: Butterworth-Heinemann.

Serial Number

C Petterd, Australian Federal Police, Canberra, Australia

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0595

Introduction

Serial numbers are present on many everyday objects. They are used to distinguish items from others of a similar appearance. In most instances, they are simply a sequential number, which describes how many units in a particular production sequence have been produced. In some cases, they are encoded to carry information about the product, such as date of manufacture, model type, etc.

Serial numbers are regularly removed from items in order to hide their true identity. The reasons for the removal are varied, and include:

- the perpetration of fraud, such as portraying an item to be of a different value when making an insurance claim;
- to hide the origin of items used in the commission of a crime, such as removing the number from a firearm used in a murder;
- to prevent positive identification of stolen property.

Serial numbers can be applied to objects by way of adhesive labels, or direct impression on to the actual object. Generally, direct application on to the object is used when the item is composed of metal, and labels are used on plastic items. Inevitably, there are exceptions to this. Cameras, for example, may carry serial numbers stamped directly into a polymer section of the body.

Methods of Application

Serial numbers can be applied in a variety of ways, depending upon the substrate to be marked and the environment the number will be used in. The major methods include:

- *Die stamping* In this technique, a male or 'inverse positive' of the character to be stamped is created, and then applied to the substrate under sudden pressure (often by hand). An indented character is left behind. This method is commonly used on metal substrates, particularly in the motor vehicle manufacturing industry. The process is often called 'cold working'.
- *Rolling* A similar process to die stamping. In this case, the characters to be stamped are applied to the substrate with a slower, more steady pressure. Its application is in cases in which the substrate could be damaged by the sudden impact used in die stamping.
- *Hot stamping* Polymer substrates are not amenable to either of the cold working methods described above. However, by heating a die prior to stamping, it can be used to produce a suitable impressed character.
- *Pin stamping* Small pins are used to form individual dots on the substrate, using an impact process. By arranging these dots into the appropriate pattern, the requisite characters are formed. The process is analogous to dot matrix computer printing.
- *Engraving* The substrate is cut away by a tiny spinning head or similar, leaving marks which form the serial number. This method is particularly common on plastic substrates, or where numbers are being applied by a consumer (for example, inscribing a driver's license number on to the back of a video cassette recorder).
- *Laser etching* A technique that is coming into widespread favor, this effectively involves burning the serial number into the substrate by applying an industrial laser.
- *Embossing* Involves a technique similar to die stamping or rolling, but is used on thin metal plates to produce a raised print appearance. The die is pushed on to the plate from behind.
- *Traditional printing methods* A variety of printing techniques can be used to apply serial numbers, especially on to adhesive labels or cloth stock. The methods include ink processes such as offset printing, hot foil stamping, screen printing, laser printing, etc.

Removal Methods

Traditionally, methods used to remove serial numbers involve a physical abrasion of the substrate until the number is no longer visible. Methods to do this include filing, grinding (usually with an electric or air-powered angle grinder), sanding with sandpaper or emery paper, or scraping with a sharp implement, especially on plastic substrates.

Numbers can also be obliterated so that they are no longer decipherable. One method of accomplishing this is called 'peening'. A relatively sharp object, such as a centre punch or cold chisel, is hammered on to the area of the number until it is damaged beyond legibility.

In the case of printed numbers, chemicals can be used to wash away or blur the numbers. Solvents such as alcohol, acetone or similar can be used.

Restoration Methods

Forensic practitioners are most often confronted with die-stamped serial numbers which have been ground off metal substrates. This is particularly common in cases involving theft of motor vehicles or the use of firearms.

If a number is ground off a surface until no visible sign of the characters remains, it does not necessarily follow that all traces of the number have been obliterated. The process of cold working involves transmitting a shock into the substrate, and compressing the metal structure such that the material which originally occupied the volume taken by the characters is displaced downwards into the substrate (**Fig. 1**). If the top layer of the substrate is subsequently removed, a damaged area of material underneath the original characters may remain (**Fig. 2**). Note that a

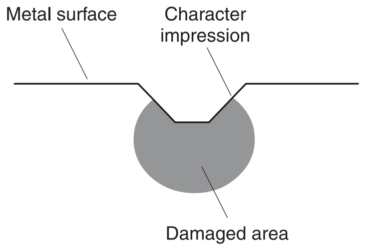


Figure 1 Damaged area underneath a stamped character.

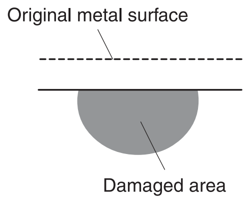


Figure 2 Residual deformation still present after obliteration of character.

number of factors may influence how far into the substrate the damaged area extends. These include the composition of the material, as well as the force used to stamp the number. If the surface is ground off deeply enough, it is possible that all traces of the damaged area will be obliterated.

There are two major methods employed for recovering obliterated serial numbers in metal substrates. These are generally referred to as 'chemical etching' and 'heat treatment'. The purpose of both processes is to create a visible contrast between the damaged and undamaged regions of the substrate. This contrast is the result of differential reflection or scattering of light from the damaged area compared with that from the undamaged area (**Fig. 3**).

Before either method can be applied, the surface must be prepared. This involves removing as much as possible in the way of residual irregularities, in order to obtain as smooth a surface as possible. A compromise must occasionally be reached when deep grooves are present, as removing the surrounding metal may achieve total obliteration of any remaining traces of the original characters.

Progressively finer grades of waterproof abrasive paper (known as 'wet and dry' paper) should be used to gradually smooth out the surface, with the best results being obtained from a mirror-like finish. Some practitioners use small hand-grinding or sanding tools to smooth the surface, but this can run the risk of uncontrolled heating of the area, as well as having less control over the depth of the abrasion. For these reasons, hand sanding is generally preferred.

Chemical etching

This is perhaps the most commonly used technique for recovering obliterated numbers. It originates from metallurgical examinations, in which etching solutions are applied to cross-sections of metal in order to observe the crystal structure under reflected light.

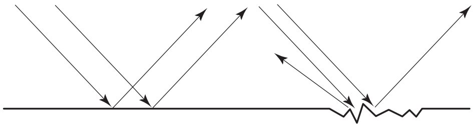


Figure 3 Reflection of light from smooth surface compared to rough surface.

For recovering obliterated numbers, the method relies upon the phenomenon that the rate of reaction between the applied chemicals and the substrate can differ in the damaged and undamaged areas. The damaged area usually has a different electrochemical potential to the undamaged surroundings, so can therefore be attacked selectively, or at least at a different rate, in comparison with the surroundings.

The chemicals used depend upon the composition of the substrate. They range from simple alkaline solutions to complicated mixtures which result in an oxidation/reduction reaction with the metal. Care must be taken when using some of the etching solutions, as they can be corrosive and/or toxic.

The solutions are generally applied by wiping them across the surface of the substrate with a cotton-wool or cloth swab. This process is repeated regularly, and the surface is observed closely between applications. Any visible characters should be noted immediately, as it is quite common for different characters to appear at different times during the restoration. It is also common for some characters to disappear shortly after becoming visible.

Another method of application, particularly useful for aluminum alloy engine block numbers on horizontal surfaces, is to form a dam of plasticene around the area, then pour the etching solution into this to a depth of 2–5 mm. The surface will be observed to bubble, with the recovered characters being visible.

A modification to the standard chemical etching technique is that of 'electropolishing'. This involves adding an electric current to the etching process. A d.c. power supply is used, with one terminal connected to the body of the substrate, and the other to the swab. This method has found favor in some forensic laboratories, but opinion seems to be divided. In many instances the extra time and equipment required are considered not to produce a superior result.

Heat treatment

This is a highly successful technique, with particular application to restoring obliterated serial numbers on cast-iron substrates. It has one significant advantage, which is speed. A heat treatment can at times be completed in a matter of minutes, whereas chemical etching is generally a much slower process. It again relies upon the residual stresses remaining under the stamped area. However, the visualization technique differs from that of etching. In this method, heat is applied directly to the obliterated area until the metal glows a light cherry red. This results in the release of the residual tensile stresses, and allows the compressed

area to bulge above the surroundings (**Fig. 4**). After heating, the area is lightly rubbed with abrasive paper, which removes any soot or oxide layer from the raised characters. They therefore stand out in contrast to the dark surroundings.

There is no specific temperature at which a heat treatment recovery will take place. Factors including the depth of the impression and the exact composition of the substrate will affect the temperature required. As a guide, the surface should be slowly heated to no more than a cherry red. An oxyacetylene welding torch with a small brazing tip, such as a number 8, is the preferred heating equipment.

It is sometimes possible to see the raised characters before the surface has reached full temperature. If this occurs, there is no requirement to heat further.

In order to avoid cracking the substrate, the area should be heated slowly, and then also allowed to cool slowly. To do this, the heat source should be either gradually turned down or slowly moved away from the surface. This avoids a rapid cooling process.

Heat treatment has been used with limited success on aluminum alloy substrates. A much lower temperature is employed, and it appears preferable to apply the heat to the substrate a short distance away from the obliterated area, and allow conduction to heat the area in question. This is known as 'indirect heating'.

For thin sheet metal areas, the application of heat treatment does not seem to produce useful results: chemical etching is the preferred technique.

Other methods

A variety of other methods for visualizing obliterated serial numbers on metal objects have been reported but are not in general use. These include the following.

Magnetic particle method In this technique, the object is magnetized and then sprayed with fine magnetic particles. The particles are attracted to regions on the surface where a crack or other damage has occurred. The method has an application in crack testing of aeronautical components, and is suitable for the recovery of serial numbers on small objects such as firearms.

The method has an advantage, in that it is nondestructive, enabling other methods to be tried subsequently if necessary. It does, however, have a significant drawback, which is in the equipment required. To magnetize the object requires either running a high amperage electric current through it, or placing the object in contact with a large magnet. Also, neither of these actions is appropriate if the object happens to be a serial number stamped on an engine still fitted into a motor vehicle.

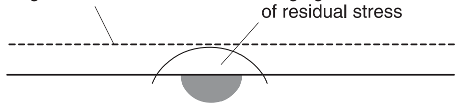


Figure 4 Bulging of surface following release of residual stress due to application of heat.

Ultrasonic cavitation Cavitation is a phenomenon that involves the formation of bubbles in a liquid due to localized reductions in pressure. The bubbles are of high energy and have the capability to etch metal surfaces.

Recovery of obliterated serial numbers in metal objects is possible if they are placed into a water bath which is excited by an ultrasonic sound frequency. It appears that the etching occurs preferentially at the sites that have been damaged by stamping. The method has the advantage of being applicable to a wide variety of metals and not requiring the use of chemicals. However, it, like the magnetic particle method, is only suitable for small objects. It is also destructive, in a similar manner to chemical etching.

X-Rays The use of radiography in efforts to recover obliterated serial numbers does not appear to be successful. It seems the radiographs are simply not sensitive enough to enable visualization of the minute damaged areas left after removal of a number. However, the method has been successfully used to locate serial numbers that have been hidden with paint or body filler, or by welding another piece of metal on top of the original.

Nonmetal substrates Less commonly encountered in casework are instances of removal of serial numbers from plastic substrates. The recovery methods described earlier are not applicable in these instances.

It appears that a stamped polymer retains a compressed area underneath the characters in a manner analogous to that of metal substrates. To recover the number, it is necessary to apply a method that will swell the substrate, allowing an image of the compressed area to become visible. The methods used to achieve this include applying heat (obviously at a considerably lower level than that applied to metal), polishing or immersing the area in chemicals.

Reagent Recipes

There are numerous formulations for chemicals that will etch different metals. For practical purposes, the major ones of interest are described in **Table 1**.

Table 1 Common metal etching formulations

Substrate	*Etchant*
Cast-iron and steel	1. Fry's reagent 90 g $CuCl_2$ 120 ml HCl 100 ml water 2. Modified Fry's reagent 15 g $CuCl_2$ 120 ml HCl 75 ml ethanol
Stainless steel	1. Acidified ferric chloride 5 g $FeCl_3$ 50 ml HCl 100 ml water
Aluminum alloys	1. Dilute sodium hydroxide 10 g NaOH 90 g water (Some sources suggest up to 60% NaOH) 2. Hume–Rothery's solution 200 g $CuCl_2$ 5 ml HCl 100 ml water 3. Dilute nitric acid 25 ml HNO_3 75 ml water 4. Alternating acid and alkali 10% NaOH and 10% HNO_3 applied alternately
Brass and copper	1. Acidified ferric chloride 19 g $FeCl_3$ 6 ml HCl 100 ml water 2. Copper chloride solution 40 g $CuCl_2$ 180 ml HCl 100 ml water

A variety of chemicals can be used to swell polymer substrates (**Table 2**).

Photography

The photographic techniques used to record recovered serial numbers are applications of those used in other forensic areas, such as fingerprints and crime scene examination.

For photographing the results of a chemical etch, black and white film has proved to be most suitable. It allows for the contrast between the faint image and the background to be enhanced. A mixture of flash photographs and available light images should be obtained where possible. Oblique lighting appears to produce the best images (**Fig. 5**). It is often necessary to experiment to determine the best angle for the incident light, depending upon the extent of the recovery, and the nature and location of the surface. A film rated at 400 ASA is suitable, in order to produce images from faint recoveries.

Table 2 Chemicals used for the restoration of obliterated serial numbers in common polymers

Substrate	*Chemical*
High density polyethylene (HDPE)	Ethanol
Polycarbonate	Ethanol; ethyl ether

The nature of the recoveries obtained by chemical etching do not always lend themselves to photography. It is possible in some instances to obtain a faint recovery which is decipherable with the naked eye but will not produce a satisfactory image on a photographic film. This reinforces the fact that careful, detailed notes should always be maintained throughout an etching process.

For recording the results of a heat treatment, color film rated about 200 ASA has been found to be appropriate. Again, a mixture of flash and available light photographs should be taken. Interestingly, the results from flash photography appear to be superior in this situation. The best results are often obtained by having the flash unit as closely aligned as possible to the lens of the camera, so that it is almost projecting vertically on to the surface (**Fig. 6**). This appears to highlight the reflective tops of the raised characters, and can result in some spectacular images.

See also: **Forgery and Fraud**: Payment Cards. **Photography and Digital Imaging**: Overview.

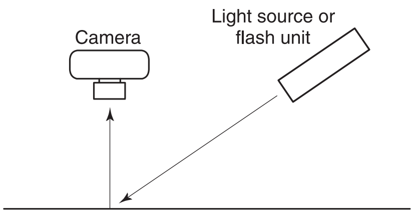


Figure 5 Oblique lighting used for photographing the results of chemical etching.

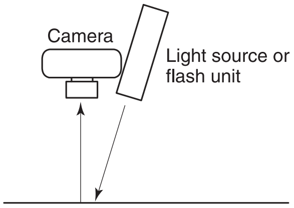


Figure 6 Lighting used for photographing the results of heat treatment.

Further Reading

Cook CW (1986) *The Automobile Theft Investigator*. Springfield, IL: Thomas.

Katterwe H (1994) The recovery of erased numbers in polymers. *Journal of the Forensic Science Society* 34:11–16.

Saferstein R (1988) *Forensic Science Handbook*. Englewood Cliffs, NJ: Prentice Hall.

Treptow RS (1978) *Handbook of Methods for the Restoration of Obliterated Serial Numbers*. Cleveland, OH: National Aeronautics and Space Administration.

Shotgun Ammunition on a Target

M S Bonfanti, University of Lausanne, Lausanne-Dorigny, Switzerland

J De Kinder, Nationaal Instituut voor Criminalistiek en Criminologie, Brussels, Belgium

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0591

Introduction

As a shotgun fires cartridges loaded with pellets, the charge emerges from the muzzle *en masse* and travels as such for about a couple of meters; then, the pellets begin to disperse, decreasing the number of pellets actually striking the target. The dispersion of the shot is clearly a function of distance, and the recognition of this relationship permits the determination of the distance between the firearm and the target at the instant of discharge. This estimation of the firing range, usually expressed from the muzzle end of the gun barrel to the target, is an essential element of the investigation. Generally, the distribution of pellets on a target allows for determination of the range of fire up to 20 or 30 m with considerable accuracy. At these distances, the shotgun pattern has no other characteristics (such as tattooing, blackening, burning, presence of a gunshot residues pattern, etc.) that could also be used for the determination of shooting distance. At closer ranges, the firing distance evaluation by shot dispersion can be complemented by analysis of the gunshot residues pattern (by optical methods and by means of specific chemical tests) and/or by exploiting others of the characteristics mentioned above.

The degree of dispersion of the shot depends on many factors, the most important of which are the type of pellets, the cartridge pressure, the wadding, the barrel length and the extent of choke. A complete analytical solution to the exact pellet dynamics is a very complicated problem, including dynamic forces, constraints and boundary conditions not completely known.

The Shot Shell

In order to fully understand various aspects of the degree of pellet spread, it is important to explain the different components of a shot shell. A shot shell consists of a cylindrical casing containing the powder (mostly smokeless), the charge (pellets or slug) and one or more wads; buffer material is sometimes present. The mouth of the cartridge is closed either by a crimp closure or an overshot wad.

The pellets (or shot)

Currently, four types of shot are available on the market: drop or soft shot (essentially pure lead), chilled or hard shot (hardened by the addition of antimony), steel shot and plated shot. Individual steel pellets weigh less than comparably sized lead pellets and thus have less range; plated shot is lead shot coated with a thin coat of copper and/or nickel to minimize distortion on firing, thus maintaining a good aerodynamic shape and increasing the range. Other softer, nonlead materials (e.g. tungsten in an organic polymer, bismuth/tin alloys, molybdenum, zinc alloys, etc.) have been tried or used recently in an attempt to increase the ballistic performance of the shot. If patterns caused by the different types of shot are examined, it appears that the harder shot gives a more uniform distribution of pellets.

Buckshot pellet patterns may be influenced at short distances by the stacking order of the pellets in the cartridge case, appearing as a series of superimposed triangles on the pattern (**Fig. 1**).

The velocity of the shot at the muzzle of the firearm is a parameter influencing the distribution of the pellets. The velocity of the shot is itself determined by the burning of the propellant powder and the length of the barrel. For traditional shot shells, the higher the pressure generated by the cartridge, the more the shot will be disrupted as it emerges from the barrel by the following gases. While this effect is largely offset by the wads used in modern cartridges, it is a factor that does affect the dispersion of shot.

Apart from the ammunition for shotguns with

Figure 1 Cross-section of a shot shell containing buckshot pellets.

smooth barrels, special rounds with shot charge exist for handguns and rifles, more particularly in calibers .22 LR, .38 Special, .44 Magnum, .45 ACP and 9 mm Luger. Flobert cartridges loaded with shot are also available.

The wadding

The wadding in traditional shotgun cartridges consists of an overpowder wad, a series of filler wads and an overshot wad. This construction suffers from a number of drawbacks, which can lead to distorted and enlarged patterns on the target. In fact, on firing, some hot gases from burning powder are able to bypass the overpowder and filler wads to reach the shot charge; these gases partially melt and fuse together a number of pellets. Likewise, the filler wads do not provide sufficient cushioning of the shot to prevent distortion due to intershot contact; in extreme cases, the rapid acceleration of the shot charge causes pellets at the bottom of the charge to be 'welded' together by the pressure into small clumps. Apart from this effect, the filler wad can fly into the rear part of the shot and scatter it.

Furthermore, during the passage of the shot through the barrel, pellets in contact with the inside surface of the bore are distorted or abraded as a result of both pressure and friction. Deformed or abraded pellets are aerodynamically less stable because they will suffer greater effects from air resistance. Consequently, they will not retain as much speed and energy, resulting in a trajectory deviating from the flight path of the main body of shot. These so-called 'fliers' are present in all shotgun patterns and can also be due to irregularly shaped shot, or pellets that become imbedded in the wads, etc. Long-shot columns aggravate the conditions described, thus leading to increased pellet deformation and hence a greater occurrence of fliers.

Another impairment to a good pattern is the overshot wad; this is supposed to slide off to one side of the shot column as it emerges from the barrel. This does not always happen, and the overshot wad sometimes remains in the shot column, disrupting it and causing a cartwheel pattern. If the overshot wad is found on the target, it is an indication of a close shot.

Modern wads tend to be of the plastic cup type with an integral shock absorber and gas seal (**Fig. 2**). The shock absorber consists of a semicollapsible section that very effectively cushions the shot column at the moment of acceleration. The integral plastic cup protects the shot during its passage through the bore and the plastic gas seal prevents the leakage of hot gases into the shot column. Cartridges loaded with this type of one piece wad (monowad) will give a much more controlled spread of shot than one loaded with the old type of wad column. The walls of the cup are split so that after exiting the air pressure divides

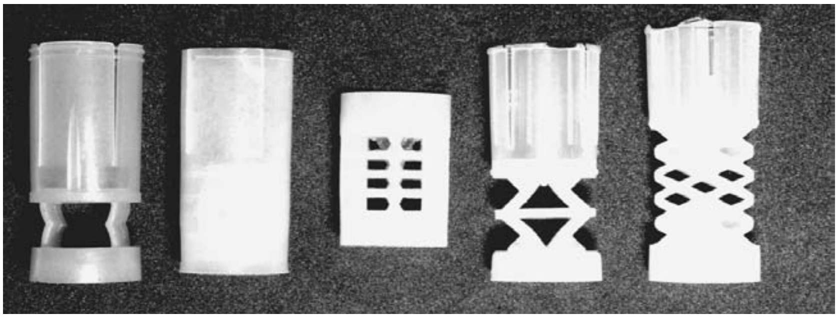


Figure 2 Different designs of plastic cup wads.

the walls into sections, resulting in more air resistance while not disturbing the trajectory of the shot. Crimping or folding of the casing material eliminates the necessity of having an overshot wad.

Shotgun wadding trajectories When a shotgun is fired, the trajectory of the different compounds of a wadding depends on its weight and shape. The wadding can be projected to 30 m or further and it can deviate from the center of the flight path of the shot. The lighter weight for a less aerodynamic shape enhances the air resistance, so that the total length of the trajectory is considerably reduced. As the wadding strike effects can sometimes be extremely reproducible, they allow the forensic scientist to conduct firing tests to determine the range of firing.

Apart from the shooting distance determination, the examination of the wad will give an indication of the caliber of the shotgun and the make of the ammunition.

The buffer material (or granulated filler)

The buffer material, mostly made of polyethylene or polypropylene and found primarily in larger birdshot and in buckshot loads, cushions the shot pellets on firing, reducing shot distortion and improving the shot pattern. The filler accompanies the shot toward the target and can produce stipple marks on the skin identical in appearance to powder tattoo marks. Marks from the filler can vary from large and irregular, to small and regular, depending on the size and shape of the granules. Buffer material can be consistently deposited on targets up to 4.5–6 m from the muzzle, and along the ground the distribution of buffer may be seen up to about 9 m. The distribution diameter of buffer also shows a linear expansion as a function of range of fire. The rate of expansion is, however, far in excess of the pellet distribution diameter, resulting in a near constant buffer distribution diameter beyond about 3 m. Crosswind can produce significant distortion in the buffer distribution, and therefore range of fire estimations based on this parameter should be approached cautiously.

The Barrel

Shortening the barrel by sawing off the end of the muzzle does have some effect on the spread of shot. What effect exists, however, is mainly due to the high-pressure gases disrupting the shot column as it exits from the barrel. Shotgun propellants are very fast burning, giving rise to a very sharp increment in pressure during the first few moments of ignition. In full-length barrels the overall pressure within the barrel drops very considerably as the shot nears the muzzle, and the volume of gas between the over-powder wad and the breech of the weapon increases. As the barrel is progressively shortened, the pressures being exerted on the base of the shot column as it exits the barrel become progressively greater. These pressures can lead to a destabilization of the shot column, resulting in a 'blown' pattern; this effect, which increases with each shortening of the barrel, can be identified by an irregular shot pattern and a larger spread than would normally be expected. For some brands of cartridges this increase in spread of the pattern is important; for other brands of ammunition, the spread remains constant.

Another effect of shortening the barrel is to remove the effect of the 'choke'.

The choke

Most shotgun barrels have a constriction applied to the muzzle end of the weapon's bore so as to control the spread of the shot. This constriction is called 'choke' and may start anywhere from 2.5 cm (1″) to 15.2 cm (6″) from the end of the barrel. The choke constricts the diameter of the shot column, increasing its overall length as well as the velocity of the pellets at the front of the shot charge. The cone-shaped reduction in the barrel results in the outer layers of shot in the column being given an inward acceleration, which delays the spreading of the shot once it leaves the barrel and reduces its tendency to separate in flight. A certain number of pellets suffer some deformation in the process, both from friction with the barrel walls and by their interaction with each other.

Some barrels intended for use at close ranges are bored without choke and are referred to as being bored 'true cylinder'; the lightest choke in the English system is referred to as 'improved cylinder' (0.12 mm), followed by 'quarter choke' (0.25 mm), 'half choke' (0.50 mm), 'three-quarter choke' (0.75 mm) and, finally, 'full choke' (1 mm). It should be noted that American and Continental choke designations are slightly different. Different degrees of choke will give different patterns for a particular shotgun charge; the tighter the degree of choke, the smaller the pattern of shot at the target. At the same range and assuming the same barrel length, cartridge type and pellet size, all bores with a given choke having different calibers (with the exception of the .410) produce approximately the same size patterns. Obviously, the patterns will differ in density.

The degree of choke is based upon the percentage of pellets that will stay inside a 76.2 cm (30") circle at 36.6 m (40 yards), with exception to the .410 shotgun, in which the pattern of shot is determined in a

50.6 cm (20") circle at 22.9 m (25 yards). **Table 1** gives the percentage of shot that can be expected for the various choke borings. There are slight variations, which depend upon the size of shot used; if cartridges of different shot size are loaded to the same average velocity the large shot will show a better result. It should be stressed that the percentages of Table 1 may be higher when using modern ammunition because of improvements in shot shell design. This improvement in pattern performance is true for all chokes and it decreases with smaller shot sizes.

In addition to the barrel choke and barrel length, the condition of the bore (either smooth or rifled, corroded, damaged, etc.), the length of chamber cone and the size of the chamber will affect the pattern. For example, if the bore of a gun has become pitted due to corrosion it will increase the friction of the shot.

Shooting Range Estimation

At ranges of within 1 m, the pellets are still travelling as a compact mass, thus the entrance hole will be a single perforation of large size (approximately 2–3 cm). Contrary to intuition, it is frequently not possible to determine accurately either bore or shot size. As the range increases, the edges of the impact will have scalloped margins. Still farther away, the pellets start to disperse, leaving scattered satellite pellets holes around the central impact. The separation and the number of these pellets will increase further with increasing distance. At distances of about 3 m, buckshot pellets cause isolated impacts.

The question of correlating range with pellet dispersion is usually approached with the assumption that the entire distribution of pellets is available for examination. In practice, the target usually reflects only a portion of the pattern and thus the range estimation is seriously limited by the size of the target. Due to the relatively small size of the average human target, this applies more particularly to victims. In this case, it is impossible to estimate the range from a measurement of the whole pattern diameter; the only measurements available are the distances between a limited number of pellet holes, and the estimation of shooting distance is scarcely reliable. Partial pellet patterns sometimes allow the examiner to position the victim within the void characterizing the pellet pattern present on the crime scene.

Where more than one shot has been fired, one should be very reticent in making any positive range determination. Another important aspect is the way the target was exposed to the charge of shot at the time of shooting. In fact, when the pellets strike the surface by an angle of incidence other than 90°, the density will appear less.

Considerable effort has been devoted to the development of empirical and mathematical models to describe the dispersion of shot from the discharge of a shotgun; many formulas have been published, but several workers have questioned their applicability. The only reliable method of determining range is to obtain the suspect weapon and a sufficient quantity of rounds of the same brand of ammunition believed to have fired the questioned pattern and then conduct a series of test shots so as to reproduce the pattern of the shooting incident. The best test ammunition is that belonging to the same batch as the cartridge of the crime. The pellet patterns can be recorded on butcher paper targets at different distances from the muzzle (**Fig. 3**). If buckshot shot shells are used, the patterns at different distances can be recorded at the same time by using an in-line array of several thin paper targets placed at known intervals from the muzzle of the shotgun. Although such an approach seems rather unsophisticated, it is capable of a remarkable degree of accuracy.

The generalizations available in the literature provide only a very rough estimate, as the variations of the type of powder, type of load, length of the barrel, degree of choke, etc. are too great. For example, some researchers have shown that two lots of ammunition manufactured within days of one another show similar ballistic performances, while those manufactured months apart have very different ballistic performances. The conditions in which shot shells have been stored may seriously compromise the determination of distance; shot shells that have been stored in cold weather may produce pellet patterns on a much more restricted range than those stored under more 'normal' conditions. In fact, the reduced temperature produces a slowing down of the burning process of the gunpowder, which reduces the pressure in the barrel chamber and lowers the velocity of the shot. Consequently, there is less natural deformation of the pellets, which results in less air resistance and smaller dispersion of the pellets, at least at close distances.

Should there not be enough rounds available, a useful range of fire estimation can be obtained by

Table 1 Percentage of shot at 36.6 m in a 76.2 cm diameter circle, according to the different chokes

Boring of gun	*Percentage of shot*
True cylinder	40
Improved cylinder	50
¼ choke	55
½ choke	60
¾ choke	65
Full choke	70

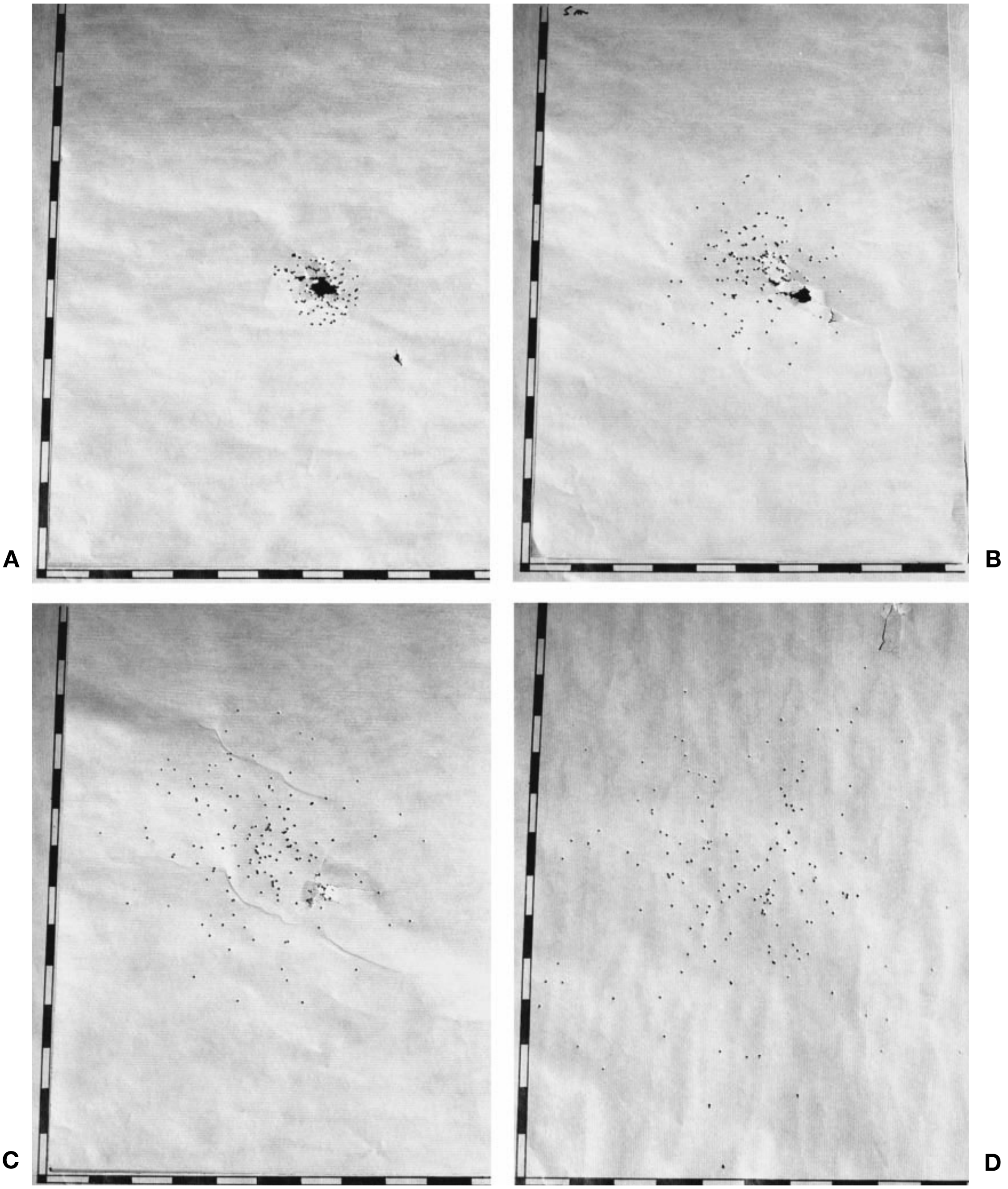


Figure 3 Shot dispersion recorded on paper sheets at various muzzle-to-target distances: **(A)** 2.5 m; **(B)** 5 m; **(C)** 7.5 m; and **(D)** 10 m. A 12/70 Mossberg 500A true cylinder riot gun was used to fire Sellier and Bellot shot shells containing 32 g steel shot with a diameter of 3.5 mm.

firing a different brand of ammunition. A scaling factor allows one to transform the results, based on a comparison of the patterns obtained with the two different ammunitions at the same distance.

The test shots on paper provide a permanent record of the effects of blackening, tattooing, burning, powdering, buffer and wadding impact marks as well as pellet spread. One has to take into account that no shotgun/shot shell combination will fire exactly the same size pattern at a given distance repeatedly, and also that some guns and loadings will pattern more uniformly and with greater regularity than others. It is therefore advisable to illustrate the variation in pattern size at a given range by carrying out a sensible number of test firings. In practice, it is suggested that at least three shots should be fired at each test distance; of course, the more shots fired at a given distance, the more accurate the findings will be.

Generally, as the distance from the muzzle to target increases, so does the amount of variation in pattern size. The larger the variation between shots, the larger the range of distances at which the shot could have been fired is going to be.

The spread of the pellet pattern can be characterized by using several parameters, based on the measurement of a characteristic length or number. The most common among them are:

1. The area or the square root of the area of the smallest rectangle that encloses the entire pellet pattern.
2. The greatest average diameter (or greatest diameter method), which is the measure of the distance between the outermost pellets in the pattern and enables the determination of the average diameter of the pattern.
3. The radius of the smallest circle enclosing the total pattern.
4. The equivalent circle diameter (ECD), which corresponds to the circumference of the polygon formed by the outermost pellet holes divided by π.
5. The distances between pellets within the pattern.
6. The effective shot dispersion $(ESD)=\{(4/N_o)\Sigma R_i^2\}^{1/2}$ where N_o is the total number of pellets and R_i is the radial distance of the i-th pellet from the center of the pattern.
7. The percentage of shots being dispersed outside a circle of a certain diameter about the pattern center.

In all these measurements, flier pellets have to be discarded; it is not uncommon for fliers to increase the actual pattern diameter to twice that of the bulk of the shot charge

The number of pellets to be accounted for will influence the choice of a particular technique. In this article, we will shortly digress on techniques (4) and (7), as we feel they are well suited for buckshot and birdshot, respectively.

For a large number of pellets, their distribution around the center should be rotationally symmetric, while the radial component *r* reflects a gaussian distribution. Recent experiments have cast doubt on its gaussian character for the central region of the distribution. Given a large number of pellets and that we are merely interested in the outward region, the probability of finding a pellet at a distance *r* to the center is given by:

$$P(r) = \sqrt{\frac{2}{\pi}}\frac{1}{\sigma}\exp(-\frac{r^2}{2\sigma^2})$$

The proportion of pellets falling within a circle with radius *r* is given by the integration from 0 to *r*. In this formula, σ is the width (mean deviation) of the distribution. The factor $\sqrt{2}/\pi \cdot 1/\Sigma$ ensures that the distribution is properly normalized when integrating of all radii $0 < r < \infty$. Hence, all the pellets can be expected to be found only within a circle having an infinite diameter. For a circle with a given radius, the incident proportion of pellets is given in **Table 2**. This table provides a means of determining by measuring the radius of the circle containing 95 or 97% of the pellets.

A statistical analysis of the results obtained by the greatest diameter method (2), as well as from the ECD method (4), shows that the ECD method gives the most reliable results. Confidence limits can be obtained by performing a statistical analysis on several test firings or by simply taking the upper and the lower limit of a limited set of values. In the latter case, the obtained confidence will depend on the number of test firings. The results can be plotted on a graph (width versus range of fire), together with the confidence levels. The pattern found on the crime scene can be converted to a range of fire with a corresponding confidence limit (maximum and minimum likely ranges of firing) using this graph or after performing a linear regression analysis. This graphical solution accounts for the lack of confidence limits available for the pattern present on the crime scene.

Intermediate targets

Shot charges may strike an intermediate target such as a pane of glass, a screen or an arm before impacting the final target. This intermediate target will increase the dispersion of the shot and a disproportionately wide spread is observed as long as the majority of pellets hit the target *en masse*. An explanation for this phenomenon may be that pellets, particularly at close ranges, are deflected by striking each other during their passage through the intermediate target, with the resulting effects possibly enhanced by the initially higher velocity. Contact or close-range wounds may create a similar effect inside the body. In cases where the pellets pass through an intermediate target, the only way to determine the range correctly is to interpose a similar target when firing. In some cases, no

Table 2 Percentage of pellets within a circle with a given radius around the center of impact

Radius	*Percentage of pellets*
σ	68
1.64σ	90
2σ	95
2.17σ	97

reliable firing range determination can be made. Heavy clothing can act as an intermediate target, enlarging the pellet pattern present on the body of the victim. For decomposed or burned bodies, it may seem convenient to obtain the pellet pattern using X-rays but experiments have shown that this method is completely unreliable.

Intermediate targets may also reduce the number of pellets in the pattern by ricocheting or stopping the pellets, as, for example, occurs when trees are in the line of fire.

Adaptations

Certain adaptations, to both gun and cartridge, can be employed to increase the shot spread. Special slow twist rifled barrels are made for such purposes; special shot spacers can be used in the cartridge loading; some devices, called shotgun diverters, can also be attached to the muzzle end of a shotgun in order to change the normal circular pattern of shot to a controlled, predictable, rectangular or ellipsoidal pattern, the notion being that the increase in lateral spread will give a higher probability of hits.

Some Misconceptions

One common misconception is that the distance, in inches, from the center of the pattern to the point where the wads hit the target gives the range in yards. This is totally untrue and should never be used for estimation of range of firing. Another misconception is that in heavy rain the pellets will be disrupted by the raindrops.

See also: **Serology**: Bloodstain Pattern Analysis. **Firearms**: Weapons, Ammunitions and Penetration; Range; Residues. **Clinical Forensic Medicine**: Evaluation of Gunshot Wounds.

Further Reading

Di Maio VJM (1985) *Gunshot Wounds: Practical Aspects of Firearms, Ballistics and Forensic Techniques*. New York: Elsevier.

Hatcher JS, Jury FJ and Weller J (1957) *Firearms Investigation, Identification and Evidence*. Harrisburg, PA: Stackpole Co.

Heard BJ (1997) *Handbook of Firearms and Forensic Ballistics*. Chichester: Wiley.

Sellier K (1991) *Shot range determination*. In: Maehly A and Williams RL (eds) *Forensic Science Progress*, vol. 6. Berlin: Springer.

Warlow TA (1996) *Firearms, the Law and Forensic Ballistics*. London: Taylor and Francis.

Tools

T M Van Dijk, South Australia Police Department, Adelaide, Australia

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0592

Overview

Toolmark examinations form part of the physical comparative disciplines within the criminalistics branch of forensic science. The primary objective is to relate one physical object with another, by determining if a toolmark was made by a particular implement. Based on the concept that the implement has transferred a design or pattern to the marked surface, simple physical comparative examination techniques are employed. A determination that the transferred pattern contains both class and/or individual characteristics, will allow the examiner to report findings ranging from exclusion through to individual identity.

Definitions

A number of terms used in this topic are explained below.

A *toolmark* is defined as that mark left by the forcible contact of any implement with a softer surface. The nature and quality of the toolmark will be determined by:

- the character of the implement itself;
- the relative hardness of the two surfaces;
- the magnitude of the force applied;
- the relative motion of one surface over the other.

Multiple contact marks are those produced by such implements as saws, files, angle grinders, or even knives used in a sawing action. The use to which these implements are put is such that their working surfaces will contact the marked material on numerous occasions. While a considerable amount of useful information can be gleaned from the examination of such marks, normally only class characteristics can be identified.

Single contact marks on the other hand are those where the implement contacts the surface once on each application. It may be that the implement is applied to the surface on several occasions, but each application is a reflection of the implement's working surface. As such, single contact toolmarks are most suitable for identification work, and if individual characteristics are identified can often be reported as having been produced by a single implement to the exclusion of others of the same class.

Only two types of mark, the striated and the

impressed toolmark, are produced by implements. Both types can be encountered in the one mark.

The *striated toolmark* is that mark left when lateral movement between the two contacting surfaces leaves a series of parallel striae (lines of ridges and valleys). (A good example of this type of mark, is a knife blade scraped across the surface of butter, leaving a contoured series of parallel lines indicative of the serrations of the knife blade.) The mark is produced as a result of a dragging or sliding motion, with minute imperfections existing on the contacting edge of the implement determining the depth and width of each striation left behind. The width, depth and number of these striae are generally quite individual and very diagnostic for the specific implement. While some implements, such as bolt cutters, axes or chisels, are designed to leave striated marks, many others can inadvertently leave such marks.

The *impressed toolmark* is produced when a hard object comes into forcible contact with the softer surface and no lateral movement occurs at the time of contact. Also described as depressed or compression marks, the features within such marks can be recognized as a negative 'mirror' image of that portion of the harder object that made contact. Such marks often include the transfer of minute (individual) imperfections present on the harder object.

Class or individual characteristics are the terms used to describe features identified within toolmarks.

Class characteristics are those features of a mark that identify it as having been produced by a specific type of implement. Often described as those provided by the original manufacturer of the implement, these class characteristics include features of size, general shape or profile, the type of cut or impression, and sometimes the direction of travel. Class characteristics are those generally retained throughout the working life of the implement.

Individual characteristics (also known as accidentals) are those features of a toolmark that are specific to the implement. Imperfections or irregularities on the surface of implements are (unintentionally) produced either at the time of manufacture or develop during subsequent use, abuse, corrosion or damage. Imperfections of this nature are random in nature and are subject to change over time. Some imperfections are large, such as when a chip of metal breaks off the tip of a screwdriver, most are microscopic, but all are unique to the implement.

Application in Criminalistics

If toolmarks are to be utilized to their full potential, investigators in general, crime scene investigators and crime laboratory personnel, all need to have a sound appreciation of the range of physical evidence to which this discipline may apply. The identification of both projectiles and cartridge cases (firearm identification) is recognized as a specialized toolmark examination discipline. The marks left by screwdrivers, crowbars or other implements used to force doors, premises, etc., axe marks in wood, the restamping of vehicle engine numbers, the marks left by pliers, bolt cutters, knives, etc., the use of hammers, multigrips and other implements, have all been the subject of toolmark identification work.

The definition of a toolmark, however, allows an even broader range of evidence types to be examined and reported on. By careful assessment, implements not normally associated with toolmark work (e.g. the marks left by teeth, the bumper bar of a vehicle or even a smear left by the edge of a shoe) could provide useful toolmark evidence. The range of impressed toolmark evidence is far broader than is often realized, and lateral thinking will often identify new applications. (The indentation left by a knee in the steel dashboard area of a vehicle allowed the crime scene investigator to nominate the person with the knee injury as the driver at the time of the accident.) (Recognizing that a heat sealer used in the production of drug deals might produce impressions on the plastic bags allowed a specific heat sealer to be identified; see Case study 6.)

Multiple contact marks Multiple contact marks, by their nature, will mask any individual characteristics, and examiners generally need to limit any identification findings to those relating to the class of implement. Examination of such marks may provide some initial assistance in certain investigations, but will generally have limited evidentiary value. (The width and sometimes the spacing of the marks left by a saw's teeth may allow the examiner to exclude specific implements.) (In one case, marks left on the blade of an axe during sharpening were such that a coarse grinder rather than a sharpening stone was indicated. While clearly not able to identify the specific grinder, the examiner was able to demonstrate that the grinder in the defendant's shed could produce marks similar to those on the axe.)

Single contact marks Almost any implement capable of leaving either an impressed or striated mark (or combinations of the two) should be considered as potentially capable of being identified. Extrusion marks produced on objects during manufacture are in fact striae left by the minute imperfections within the extruding tool. Such marks can be found on copper wiring, extruded plastics and even aluminum foil, and may be used to identify the origin of such

evidence when located at a scene. (The extrusion marks on a plastic rubbish bag containing body parts were favourably compared with a carton of such bags in the possession of the defendant.)

To best illustrate the broad range of evidence types that lend themselves to toolmark identification the following illustrations, in the form of case studies, are provided.

Case study 1

The extrusion marks on the inner surface of flexible irrigation piping (used at a drug crop) was identified as originating from one end of a bulk supply at a local hardware store. The extrusion pattern consistently altered within a 10 m length (presumably at manufacture due to wear and dirt developing around the extrusion gate). This evidence corroborated statements identifying the suspect as having bought the piping, thus implicating him in the establishment of the drug crop (**Fig. 1**).

Case study 2

The gate to a car-wrecking yard was found to have had the chain cut. A suspect walking some 200 m from the gate was seen to place a pair of bolt cutters in the back of a utility vehicle as police drove past. A test cut made with the bolt cutters (in lead rod) showed a replication of the microscopic contours (striae) left on the cut chain link, sufficient to exclude any other implement (**Fig. 2**).

Case study 3

Several vehicles parked along a footpath had their tires damaged with a single stab into the tire sidewalls. A penknife was recovered from a young suspect, and test cuts produced in another inflated tire were matched to one of the stab marks. The striae in this case were very fine and compressed together (due to the angle at which the knife entered the tire wall) but clearly produced a unique and reproducible contour pattern. Duplicating the original angle and direction of travel was vital before a match could be demonstrated (**Fig. 3**).

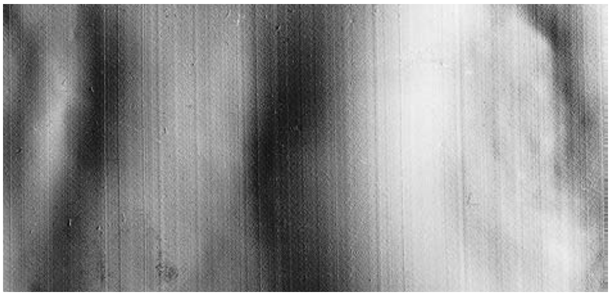


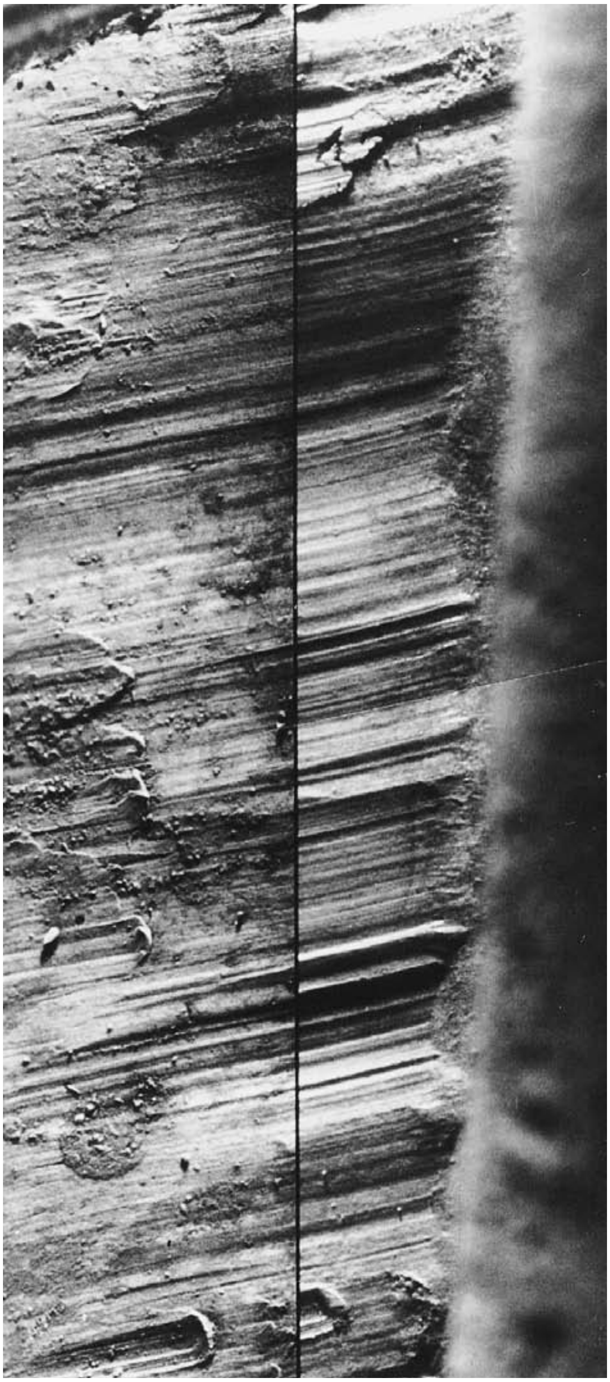
Figure 1 Case study 1. A cast of the inner surface of extruded irrigation piping. Extruded striae are quite common on a wide range of materials, including plastic bags, copper wiring, etc., and may allow a relationship to be demonstrated between two sites (objects). The major concern with such evidence relates to the length over which such comparable striae are left. A study of the manufacturing process and the quality controls in place, as well as samples from the manufacturer, should always form part of any such identification work.

Case study 4

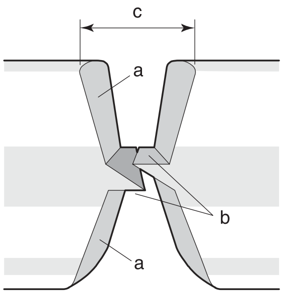
The names stamped on the edge of a number of World War I medallions were suspected of having been forged. Sets of such medallions (in the name of specific war veterans) were being sold at auction as valuable collector's items. The suspect was a jeweller who had three sets of very fine steel marking stamps. Test stampings produced in copper were compared with corresponding letters on the medallions. The use of the scanning electron microscope in this case allowed extremely detailed surface irregularities, which existed on the original stamps and which had been transferred to the medallions, to be used as a positive identification (**Fig. 4**).

Case study 5

Numerous small particles of steel recovered from the scene of a fatal parcel bombing were identified as belonging to a food can. Some of these fragments were recognized as forming part of the upper rim of the can, and minute, regularly spaced indentations along the side of the rim were identified as most likely to have been caused by the feeder wheel of an electric can opener. The marks exhibited a combination of 'impressed' and 'striated' detail created by the action of each individual tooth on the wheel digging in and pushing the can forward past the cutting wheel. Although 80% of these teeth marks were damaged by the close proximity of the blast (approximately 2 m of detonating cord had been wound tightly into the can), about 16 individual marks showed comparative detail. All 24 teeth on a feeder wheel of the electric can opener, recovered from a suspect's home, left marks which could be clearly distinguished, one from the other. Requiring the development of new examination techniques (i.e. cleansing and preserving the 'scene' marks as well as producing undamaged 'test' impressions), subsequent comparison work identified that nine of the feeder wheel teeth could be compared with the 16 useable marks located on the bomb fragments. The subject of considerable research, the identification work confirmed that it would be highly unlikely for another feeder wheel to leave marks that would correspond in both detail and position to each other. The marks shown are two of the nine teeth marks which were compared (using a standard reflective light comparison microscope) and, although the



A



B

Figure 2 Case study 2. Bolt cutter identification. **(A)** Left, scene (chain link); right, test cut in lead. A classic striated toolmark identification in which only a short segment of the cutting jaws were responsible. Identifications of this type can readily be applied to numerous similar implements such as pliers or wire cutters. **(B)** Cross-section sketch of a cut produced by bolt cutters. Although generally producing striated marks (a), examiners should also be aware that identifiable impressed detail can be left at the very base of the cut (b). When hardened steel (padlock shanks) are cut, poor or no striated features often result, leaving the impressed detail as the only possible means of identification. (c) Profile of cutter blades.

scene marks have some distortion (due to blast effects on the metal), the distinct differences of one tooth to another can be readily seen (**Figs. 5–7**).

Case study 6

The examination of detail left on a heat-sealed edge of a small heroin deal allowed a specific heat sealer to be identified. The seals on the clear plastic bags needed to be cast on both sides before the 'impressed' marks could be examined in detail. Even when new, heat sealers leave distinctive features, and numerous drug trafficking operations have since been the subject of this type of identification (**Figs 8–10**).

Many more case studies can be provided by toolmark examiners around the world, all of them involving different implements, but all based on the presence of identifiable individual characteristics within either striated or impressed toolmarks. It is the recognition of what may constitute a toolmark that will lead to an even broader use of this discipline in criminal investigations.

Limitations to toolmark evidence

Toolmark evidence is generally used to demonstrate an association between the scene of a crime (either the place or the victim) and the suspect, by means of the implement involved. It is very important therefore that evidence exists that will place the implement in the possession of the suspect at the time of the alleged offence (DNA? Fingerprints? Observations, Witnesses, Admissions?). Investigators need to be made aware of this limitation, as toolmark identification work may be waste of time and effort if such associative evidence is not forthcoming. (A screwdriver discarded at one scene may be capable of being connected to another scene nearby, but by itself has little evidentiary value in identifying the culprit.)

The working edges of many implements are subject to corrosion, wear and abuse. As such, the individual characteristics (on the implement itself), on which an individualization must be based, can be destroyed shortly after the scene impression is deposited. (In

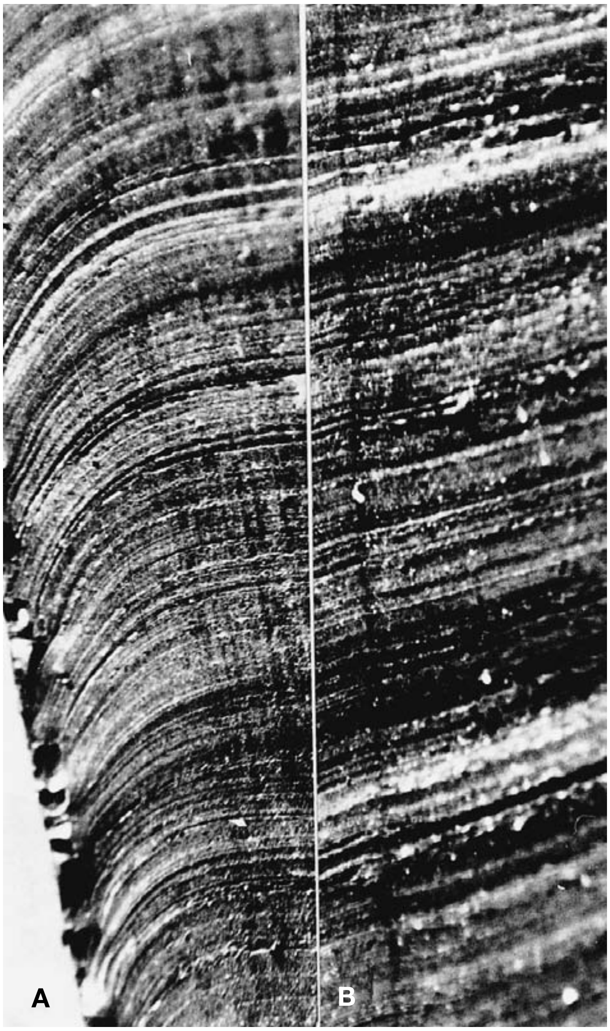


Figure 3 Case study 3. Penknife identified as having stabbed a vehicle tire. **(A)** Stab cut in tire; **(B)** test cut. Fine imperfections on the blade (or more specifically the contacting surface) of any implement will determine the number and depth of individual striae, but their spacing relative to each other is governed by the direction of travel. Thus if this penknife blade were scraped across a pat of butter, the same striations would be evident but would bear little resemblance to the mark in the tire (see also Figs 17 and 18).

fact, even as the implement is being used, damage may be inflicted.) Unlike fingerprint and DNA evidence, therefore, toolmark evidence has limited classification value and also usually has a limitation of time. Many crime laboratories discard their unidentified toolmark evidence (from breaking scenes and other volume crime type) after 6 months.

Collection of Toolmark Evidence

As with almost all forensic science work, toolmark identification relies on sound crime scene investigation practices. Toolmark evidence is more prevalent than is normally accepted, but may often be overlooked because:

- No suspect tool is available for comparison.
- The scene examiner/investigator is not aware of the potential for such evidence, or, if aware, has excluded potential toolmark evidence without proper assessment (i.e. the mark is considered too damaged or too small (see Case study 7), when microscopic examinations can often reveal far more detail than is evident at the scene).
- Investigation of suspect homes, persons or vehicles does not include a search for comparative implements/material.

Rules for collection of toolmark evidence

The following should be adhered to, whether collecting toolmark evidence *in situ* or by casting.

- General, midrange and close-up photography should be done at a minimum.
- Close-up (macro) photography should incorporate oblique lighting from several directions.
- Trace evidence (paint, metal smears, etc.) may have transferred during the production of the mark. Examiners need to record and prevent both its loss and contamination, on both the mark itself and the suspect implement. The analysis of trace material may provide further associative evidence between the mark and the implement, especially if the comparative results fall short of individualization (see Case study 8).
- *All* available toolmarks need to be collected. The field technician cannot be certain as to which mark will provide definitive detail; if not collected, valuable information, not able to be confirmed in the laboratory, may be lost. While many scene investigators will collect the striated marks left by a pinch bar, the fulcrum point may also provide valuable detail, particularly when trying to exclude implements in the initial examination stages (see Case study 7.) Both cut ends of a bolt cutter mark, when realigned, will provide a profile of the cutting blades, which often reduces the number of implements requiring detailed examination.)

Case study 7

A chemist shop's front aluminum door frame had simultaneously been levered some 30 cm above and below the central door lock, just enough to allow the door to be sprung. Minimal damage was done to the door or frame, both of which were still serviceable. While no marks were evident on the edge of the door, the two leverage points were obvious on the thin (2 mm wide) aluminum edge of the door frame. The

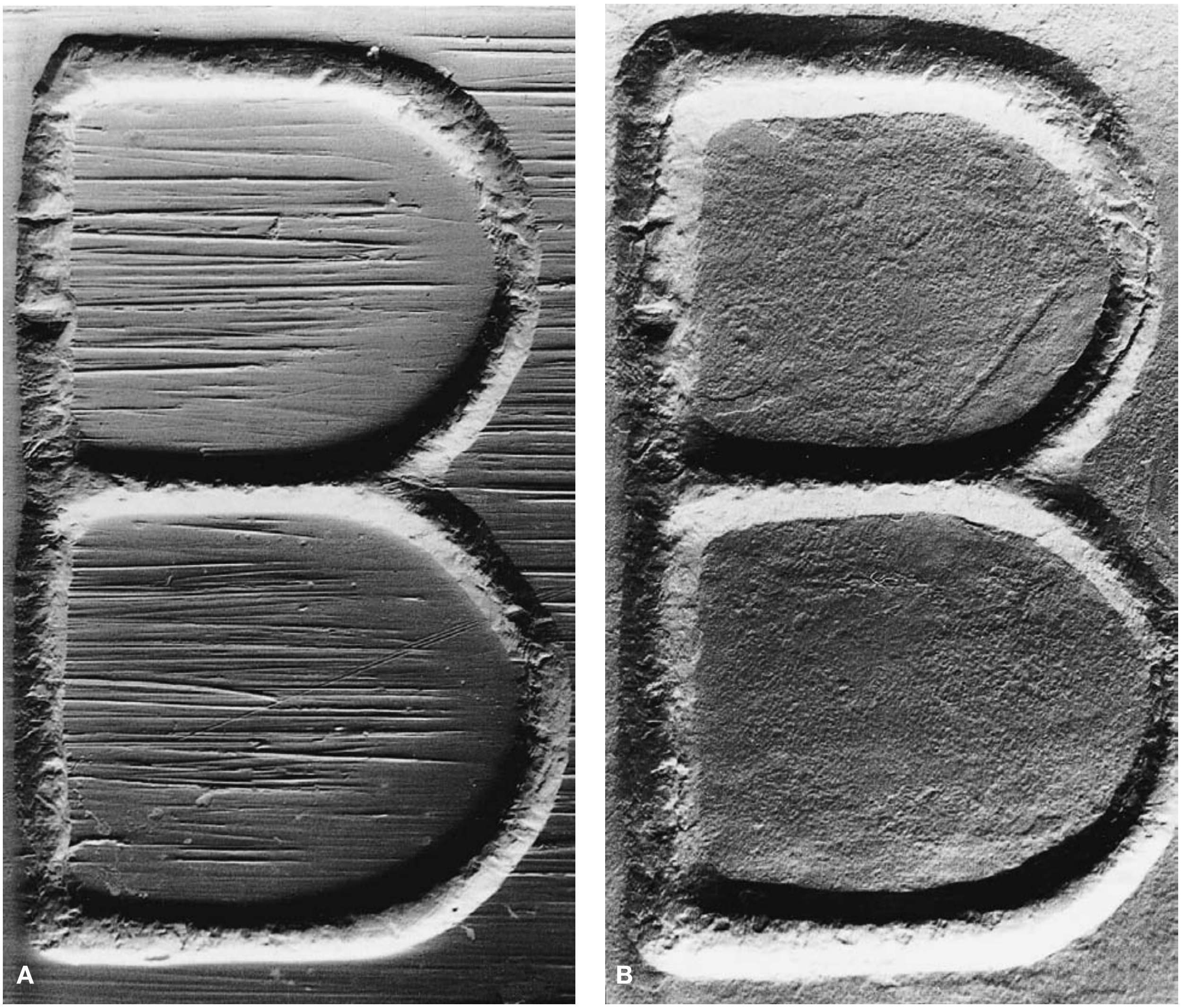


Figure 4 Case study 4. **(A)** Letter 'B' stamped on edge of war medallion; note the grind marks in the background. **(B)** Letter 'B' stamped on copper sheet using one of the jeweller's stamps. Although some differences can be seen (due either to the substrate on which it is stamped or further damage to the stamp since its original use), an overwhelming number of corresponding defects clearly identify both as having been produced by the same stamp. Standard reflective light microscopy was originally used to satisfy the examiner of the identification, but the use of the electron microscope in this case was an ideal medium for the presentation of the results. The use of this instrument should be assessed as an option whenever very small marks are encountered (e.g. firing-pin impressions on cartridge cases).

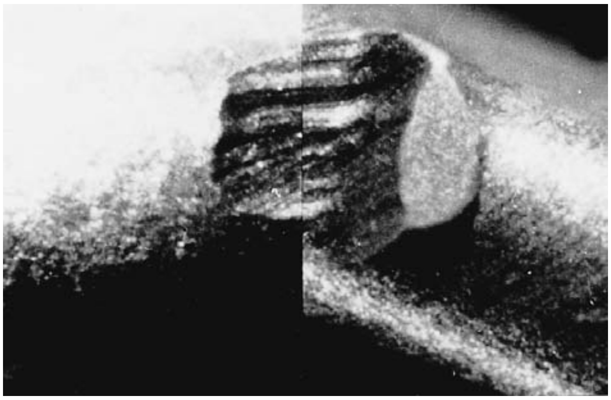


Figure 5 Case study 5. Comparison between a scene mark (left) and that left by tooth number '5' on the feeder wheel (right).

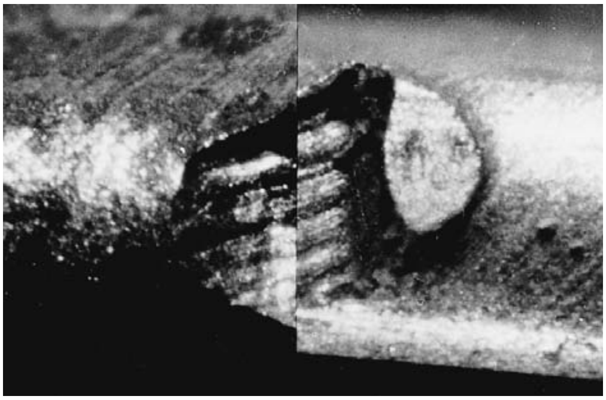


Figure 6 Comparison between scene mark (left) and that left by tooth number '1' on the feeder wheel (right).

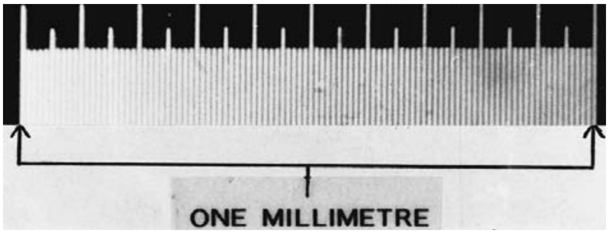


Figure 7 A 1 mm scale included here to give some indication of size. While these marks were only one-tenth the size of the marks in Case study 4 (the letter stamps), electron microscopy was found quite unsuitable for demonstrating the striae comparison results.

Figure 8 A cast of one small section of the seal produced on a plastic bag.

Figure 9 A different section of the cast shown in Fig. 8, further along the heat-sealed plastic bag. The silicon arm, which is pressed down on to the plastic bag during the heat-sealing process, has numerous minute defects whose spatial orientation produces a unique pattern, readily recognizable on any surface (or a cast thereof) that has been sealed by it.

marks, however, were quite small and showed no class characteristics. Initially not considered to be of any value, the crime scene investigator cast both the marks when advised that two suspects had been questioned and implements had been seized. As the casts were a positive replica of (albeit a very small portion of) the implements used, direct comparison between the casts and the implements themselves was adopted (i.e. no test impressions). The ability to relate both implements and the method employed in their use resulted in both offenders pleading guilty at the time of arraignment (**Figs 11–14**.)

Figure 10 Cast of the sealed surface on the opposing side of the plastic bags sealed with the heat sealer (Figs 8 and 9). The Teflon-coated cloth material which covers the heating element (on the base of such heat sealers) will also be quite different from one sealer to another, with defects developing during use.

In situ collection

The most suitable method for the collection of toolmark evidence is to collect and preserve the item on which the mark is located. Precautions may need to be taken to prevent deterioration of the mark through damage or oxidation (e.g. a steel chain link may need a thin film of oil to prevent rusting), but the ability to

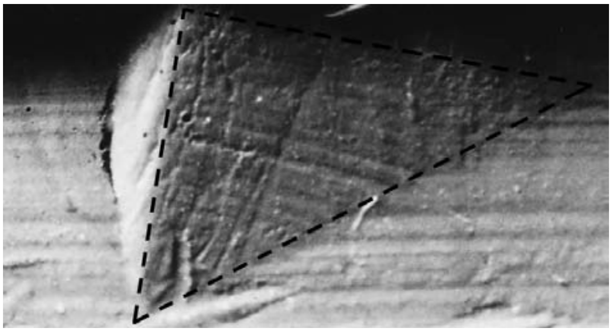


Figure 11 Cast of the lower fulcrum mark left on a 2 mm wide aluminum door-frame edge. The cast has faithfully reproduced the mark as well as some extrusion marks on the aluminum itself.

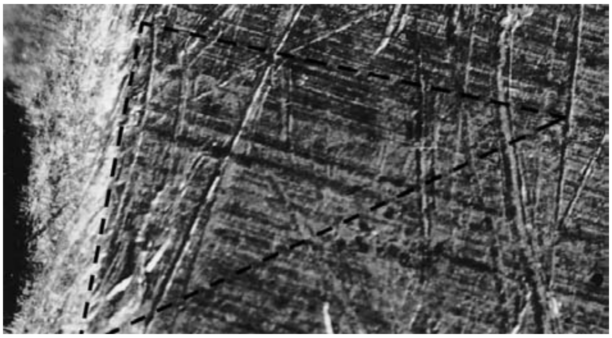


Figure 12 A photograph of a portion of one side of a small screwdriver recovered from one suspect. The detail is located some 12 mm from the actual screwdriver blade and shows both machining and wear features (see Fig. 11).

Figure 13 A second cast collected from the scene (Fig. 11) was initially discarded, because of what was perceived to be poor detail. The unusual pattern seen in this second scene cast, however, could be related directly back to a specific section on the second implement, a drop forged crowbar.

Figure 14 A small section of crowbar identified by a thin smear of aluminum. In this location scaling and other individual features left on the bar during its manufacture were directly comparable, even though a slight dragging of the crowbar across the surface had occurred (see Fig. 13).

examine the mark itself under laboratory conditions is the ideal. Many items on which the toolmark exists have been damaged in such a way that they will require replacement. As such, most victims have little hesitation in allowing the examiner to collect such evidence. Objects such as doorknobs, chain links, metal sheeting, etc., and even splintered wooden door frames, can be dealt with in this way.

Casting

An alternative to *in situ* collection of toolmark evidence is casting. Often toolmarks exist on items that do not lend themselves to collection. Altered engine numbers cannot be readily examined under the microscope. Damage to some items may be such that they are still serviceable and *in situ* collection is not an option (see Case study 7).

The most suitable casting materials are those silicon-based media that are simple to mix, easy to apply and to cure with minimum distortion. More importantly, total opacity and the ability to replicate faithfully the finest detail are the two properties that tend to reduce the options available for the collection of toolmark casts. The slightest translucence in the casting medium will reduce the visibility of the surface detail, effectively minimizing the possibilities of demonstrating an identification. Whilst most silicon casting mediums will record the detail, only a few have the opacity suitable for toolmark work.

Good casts should have few or no air bubbles, as even small air pockets can result in loss of valuable microscopic detail. The collector's detailed notes and photographs, relating the correct position and orientation of the casts, will greatly assist the examiner in determining angles and direction of application, thus simplifying comparison work. This becomes of particular value when the examiner needs to produce test impressions, which require duplication of the direction and angles applicable at the scene.

Examination and Identification

Toolmark identification work requires a simple methodology in which the evidence is analysed, compared and finally evaluated. The techniques applied are not unlike those applied in shoemark identification work, albeit normally at a microscopic level. In almost all cases the use of reflected oblique lighting, a stereo binocular low-magnification ($5-50\times$) microscope, a low magnification ($10-50\times$) comparison microscope and an image recording facility (photographic or video) will be the major requirements. Access to a simple workshop to enable the production of test impressions is also recommended.

Scene impressions

Examination of the scene (also referred to as the unknown or the questioned mark) impression is normally carried out first, to determine if it is suitable for comparison purposes. Lighting is quite critical in such examinations, and either the impression or the lighting needs to be angled to produce maximum detail. On some occasions, several lighting angles need to be employed before all the details evident in the mark can be recorded.

Such factors as the nature and quality of the mark, any class characteristics, damage sustained and the existence of individual characteristics are noted prior to any comparative work. At this time the examiner needs to identify those features that have been produced by the implement and are not the product of the substrate on which the mark exists (e.g. wood grain, extrusion or grind marks). Trace evidence also needs to be correctly dealt with before the start of comparative work (i.e. although such trace evidence may not require analysis at this time, it should be noted, recorded, separated and preserved).

Case study 8

Fig. 15 is a mark left (in the plastic coating of an exterior alarm wire), by the side cutting portion of a set of combination pliers. Although comparable to the marks left by the suspect pliers, detail in material such as this is often of poor quality and by itself not sufficient for the total exclusion of other, similar pliers. Caught in the cutting blades of the pliers, however, were copper smears and three different colored particles of plastic, similar to the insulated wiring encased within the alarm wire. The analysis of the plastic and the toolmark comparison work corroborated each other to strengthen the identification (Fig. 15 and **Fig. 16**).

Suspect implements

The suspect implement(s) are examined next. If the initial scene impressions have been well documented, much elimination work can be achieved at this time. Implements not meeting the dimensions of the scene, having an action not suitable to the production of the scene, etc. can all be excluded. Again, the presence of trace material must be dealt with before the production of test impressions. The position of such trace evidence may also assist the examiner in identifying the specific area of the implement to focus on. (The shape of the chain link can often be seen smeared on the surface of the bolt cutter blades. Without the light smear on the crowbar in Case study 7, the likelihood of locating the impressed detail would have been remote.)

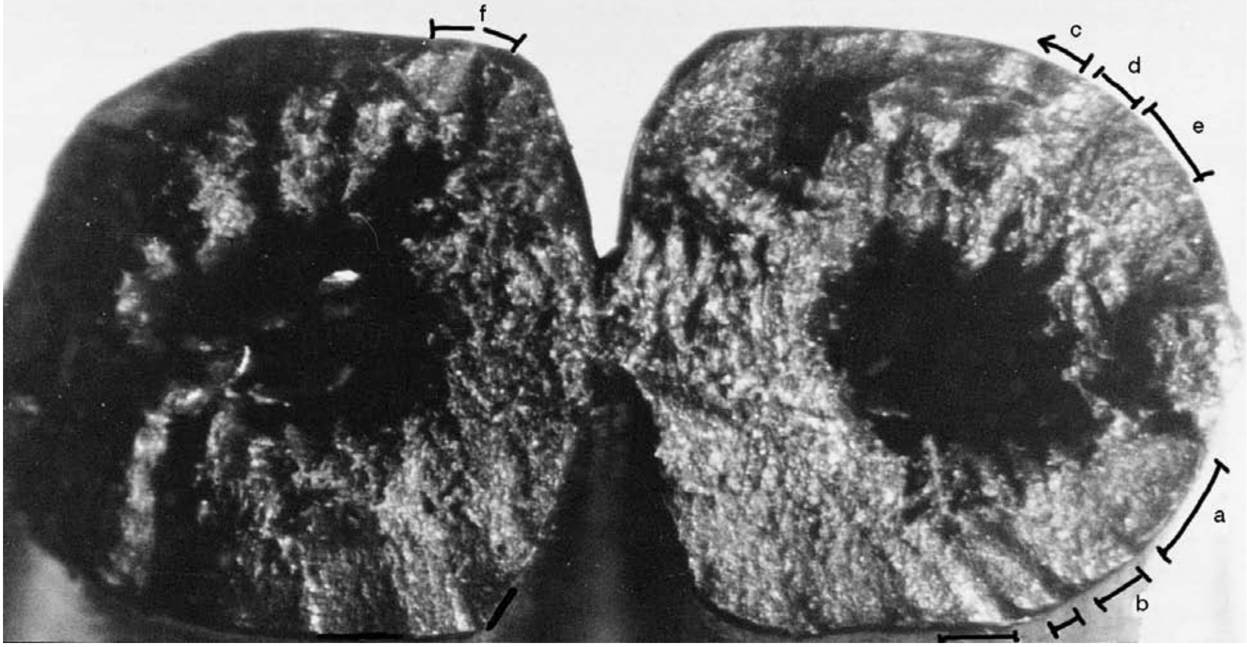


Figure 15 Cross-section of insulated alarm wires believed to have been cut with the side-cutters portion of a pair of pliers. The profile of the cut was quite distinctive and capable of being reproduced by the pliers in question but not by other, similar pliers.

Comparison and test impressions

Comparison between the scene and the implement will normally require the production of test impressions. While some impressed toolmarks (particularly casts) are capable of being directly compared with the implement, the striated toolmark can only be compared with a striated mark produced by the suspect implement. Test impressions also serve the important purpose of demonstrating that the implement is, in fact, capable of producing the mark being investigated.

Test impressions Test impressions should duplicate the conditions under which the original marks were produced, while insuring no damage is inflicted to the surface of the implement. Lead or annealed copper sheeting, aluminum or other soft-metal media are often selected, if the scene impressions were originally produced in metal. There are occasions where the implement will only produce comparable marks in the medium of the original. (This is often the case with marks produced by bolt cutters in hardened steel.) On such occasions, the initial test impressions (in the softer mediums) will provide some guidance before the final test impressions are made in the harder scene material.

Drag or scraping type marks left by crowbars, screwdrivers and the like, are the most difficult type of test impression to duplicate. Marks made by such implements can alter dramatically if the angle of application or the direction of travel across the surface is altered. Changing the angle of application will result in different defects on the implement now being

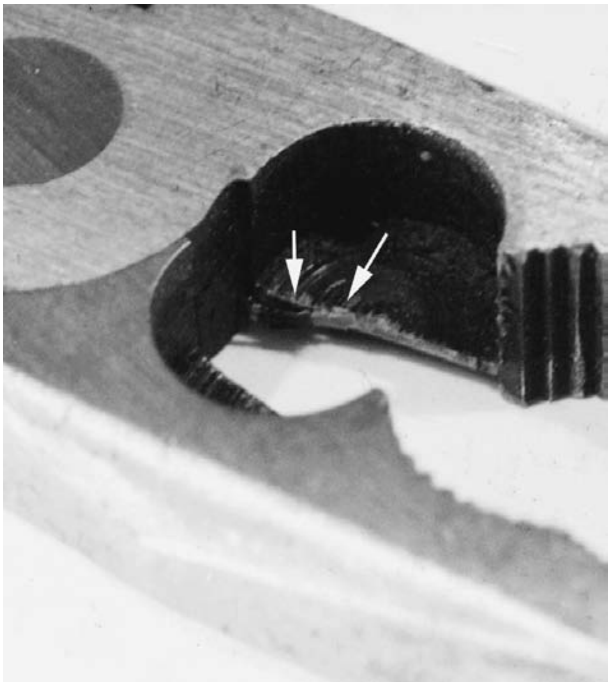


Figure 16 (see color plate 44) Macrophotograph showing multicolored plastic insulation material caught in cutter blades of suspect pliers.

in contact with the marked surface, leaving a different set of striae (**Fig. 17**). Altering the direction of travel will result in a change in the width of individual striae (**Fig. 18**). For these reasons striation marks produced by such implements can be difficult to reproduce, and will normally provide the examiner with considerable challenge. Being able to view the striated crowbar or screwdriver mark at the scene will often assist the toolmark examiner in identifying the angle of application and direction of travel. Test impressions made under similar conditions to the scene will maximize the likelihood of producing marks suitable for comparison.

Sufficient test impressions should be made to demonstrate reproducibility of the features being compared. Most examiners compare test impressions (to demonstrate this reproducibility) before comparison with the scene.

Comparative examinations Comparison between the tests and the scene can now commence. Although the comparison microscope will feature as an important element in this phase, much of the comparison work can, in fact, be performed under low magnification utilizing the stereomicroscope. At this stage the examiner locates features or groups of features that will provide a pattern recognizable in both scene and test. The shape, size and spatial relationship of such features can be quickly observed and noted. In regard to impressed toolmarks, the comparison can often be completed using only the stereomicroscope, as long as lighting conditions remain constant. For striated toolmarks, a careful examination of each mark under similar lighting conditions (particularly direction and angle) will, on most occasions, identify promising areas with similar detail.

The comparison microscope is still the simplest method of comparing striated toolmarks. The lighting conditions on both objectives must be carefully matched before a detailed search for matching contour features is possible. Angles at which the striated

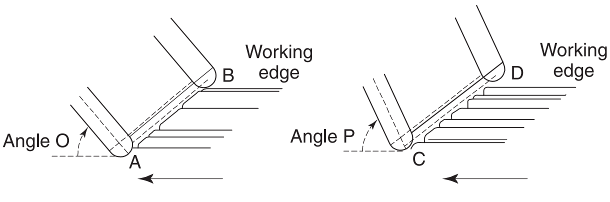


Figure 17 The Effect of alteration in an implement's angle across a surface. Since varying defects exist on different portions of an implement's blade, an alteration in the vertical angle as the implement is dragged across the surface will produce a variation in the contour features of the resultant toolmark. Reproduced with permission from Van Dijk (1993). Reprinted from *Forensic Science International* (1981) Vol 5, pp 18, Maheshwari, Influence of vertical... with permission from Elsevier Science.

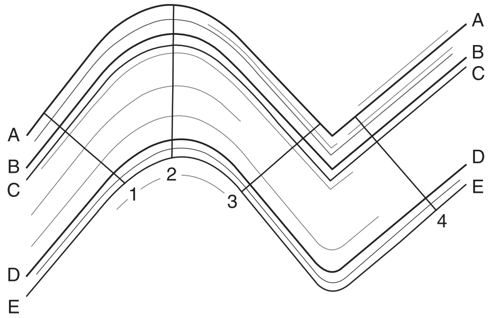


Figure 18 Effect of alteration in direction of an implement across a surface. An alteration in the lateral movement of the implement across the marked surfaces will alter the distance between corresponding striations. If locations 2 and 3 existed as two separate marks, trying to demonstrate a relationship between them would be difficult. The diagram demonstrates the need for duplicating the scene conditions to ensure that a comparison can be achieved. Reproduced with permission from Van Dijk (1993). This material has been reproduced with the express permission of LBC Information Service (a part of Thomson Legal & Regulatory Group Asia Pacific Limited). Source: *Expert Evidence* (1993) by Freckleton & Selby © LBC Information Services.

detail meets may need slight alteration, but when a match is achieved the similarities of contour are usually striking.

Impressed toolmarks can also be compared using the comparison microscope. Impressed features can be superimposed or compared side by side, confirming size as well spatial relationships.

During any physical comparative process, differences between the scene and test impressions will be inevitable. The use of different materials, damage to the scene or the implement since the mark was first produced and other factors may contribute to such differences. The examiner needs to address any significant differences, especially those that cannot be explained. Unexplained differences will affect the evaluation of any results.

Evaluation of results

As with any physical comparative work, demonstrating a relationship between the scene and the implement under investigation is based on the identification of class and individual characteristics. The existence of comparable class features only will allow the implement to be identified as one of a specific group capable of having made the mark. Whether that group (or class) is limited or extensive will determine whether such findings have any evidentiary value.

The identification of sufficient individual characteristics in both the scene and the implement determines whether individualization, to the exclusion of any other implement, can be reported. The nature of physical comparative analysis is such, however, that there is no minimum number of characteristics

required before a 'positive' can be reported. Rather the nature and shape of identified characteristics, their relative position to each other, similarity in profile or proportion of matching striae, all influence the examiner. The possibility of another implement having the identified individual features in the corresponding spatial orientation is the basis on which an identification is made. While much work on probability has been carried out, the diversity of implements, the type of marks they may leave, as well as the ever-changing implement surface, have, to date, limited the extent to which statistical data can be developed or applied. As such, possible or probable identifications should not normally be expressed in mathematical terms.

Any evaluation of identified matching characteristics is based on an examiner's knowledge of the particular implement's manufacture as well as an appreciation of the physical properties of metals. Research is often required before an examiner is able to provide definitive results as to the origin and cause of comparative features. The research undertaken on specific implements and reported on by various examiners can often be applied to implements whose manufacture incorporates similar processes.

Presentation of results

As with all forensic disciplines, the reporting of toolmark identification work needs to be unambiguous. The basis of, and any limitations to, the reported findings must be clearly expressed. Reporting in such a manner will insure that the appropriate weight can be placed on the evidence in the context of the case being presented. Two examples of findings are provided:

- The screwdriver (X) is identified as having produced the mark (Y) located on the steel filing cabinet. Based on the matching of striated detail on both Y and a test impression produced by X, this finding excludes any other screwdrivers.
- The screwdriver (X) is capable of having produced the mark (Y) located on the steel filing cabinet. Based on the presence of detail identifying size and shape, but Y lacking sufficient individual characteristics, this finding cannot exclude other flat-bladed implements (i.e. screwdrivers) of the same size and shape.

Having produced a report, it is up to the examiner to provide evidence of the stated results when called upon to do so. Clearly, notes and sketches may be of value but, as can be seen from the case study illustrations, one of the most effective demonstrations of the results achieved is provided through photographs and comparison charts. While such charts may have some limitations, these can readily be explained to any lay person. The adage that 'a picture speaks a thousand words' can certainly be applied in physical comparative examinations, and (in the experience of the author) simplifies the reporting process.

Toolmark Identification in the Future

Toolmark identification, it appears, will remain a simple physical comparative discipline for the foreseeable future. The development of improved imaging techniques and the application of computer technology have already provided firearm identification search programmes. Some jurisdictions have applied this technology to develop a database for all striated toolmarks, against which any new case can be searched. These developments will provide more objective data, which the examiner can apply in the evaluation process, but the discipline will remain one of pattern recognition and direct comparison.

Summary

Forming part of a broad range of physical comparison disciplines within forensic science, toolmark identification compares the features present in a scene mark with those present on either the implement itself or a test impression made by it. Relying on sound crime scene investigation techniques for their location and collection, as well as recognition of their value, results of any identification will also be determined by the quality and nature of the mark being investigated.

The working edge or surface of any implement will normally contain features that can be identified as class or individual. Any identification will be based on the existence of such features in the scene mark, and the ability to relate them to the specific implement in question. The application of scientific method, training, experience and research will allow examiners to identify and distinguish between the two types of identifying features. The presence of sufficient individual features will enable examiners to report identity or a relationship between the scene impression and the implement, to the exclusion of any other. The lack of such detail will either exclude the implement altogether or allow it to be placed in a class capable of having produced the mark in question.

Examinations are normally performed utilizing oblique lighting, a low-powered stereomicroscope and a standard low-powered comparison microscope. Lighting and magnification must be duplicated for both scene and known samples. Using direct physical comparison techniques, including comparative measurements, side-by-side comparison and overlays,

results of any examination can best be illustrated using comparative micro-and macrophotographs.

See also: **Crime-scene Investigation and Examination**: Suspicious Deaths. **Criminalistics**. **Evidence**: Classification. **Pattern Evidence**: 'Plastic' Bag Striations; Vehicle Tire Marks and Tire Track Measurement.

Further Reading

Burd DQ and Gilmore A (1968) Individual and class characteristics of tools. *Journal of Forensic Sciences* **13**:390–396.

Burd DQ and Kirk P (1942) Toolmarks: factors involved in their comparison and use as evidence. *Journal of Criminal Law and Criminology*, 679–686.

Butcher SJ (1975) A study of the marks made by boltcutters. *Journal of the Forensic Science Society* **15**:115–126.

Cassidy FH (1980) Examination of toolmarks from sequentially manufactured tongue and groove pliers. *Journal of Forensic Sciences* **25**:796–810.

Davis JE (1958) *An Introduction to Toolmarks, Firearms and the Striagraph*. Springfield, IL: Thomas.

Denton S (1981) Extrusion marks on polythene film. *Journal of the Forensic Science Society* **21**:259–262.

Tuthill H (1994) *Individualisation, Principles and Procedures in Criminalistics*. Salem, OR: Lightning Powder Co.

Van Dijk TM (1985) Steel marking stamps: their individuality at the time of manufacture. *Journal of the Forensic Science Society* **25**:243–253.

Van Dijk TM (1993) Toolmark identification. In: Freckelton IR and Selby H (eds) *Expert Evidence*, chap. 85. North Ryde, NSW: Law Book Co.

Warren G (1991) New system for toolmark identification: electronic imaging, recording and comparison. *Association of Firearm and Toolmark Examiners' Journal* **23**:990–996.

Vehicle Tire Marks and Tire Track Measurement

W J Bodziak, Forensic Consultant Services, Jacksonville, Florida, USA

Copyright © 2000 Academic Press

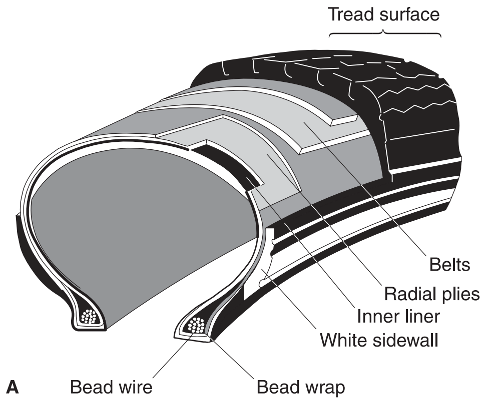
doi:10.1006/rwfs.2000.0594

Introduction

In today's highly mobile society, criminals often use vehicles to travel to and from the scenes of the crimes they commit. On many occasions, their vehicles track over surfaces which retain the impressions of one or more of the tires, leaving proof of their vehicles' presence. These impressions range from a partial impression to a full set of tracks representing all four tires. Proper recovery of the impressions of track evidence, through crime scene drawing measurements, photography and casting, serves as a way of preserving the evidence for subsequent comparison with the tires and dimensions of a suspect vehicle.

Tire marks or tire impressions are those left by the tread design of a tire after they pass over a surface. As with any mark or impression, the detail retained by the surface ranges from exceptional to poor, depending on a variety of factors such as substrate qualities like texture and color, the amount and type of residue on the tire, moisture, etc.

Tire tracks are the marks or impressions left by the tires of a vehicle, exclusive of the tread design, and include track width, wheelbase, turning radius and the relative positions of the tracks of all four tires. **Figure 1** gives some basic nomenclature.



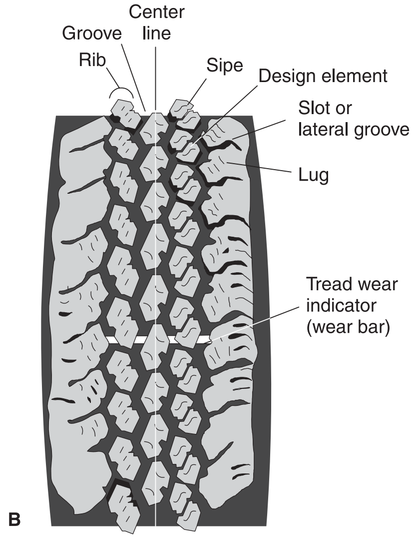


Figure 1(A) and **(B)** Radial tire: basic terminology.

Sidewall Information

Much information is molded into the sidewall. Portions of that information are of importance to the investigator and examiner, and should be noted when investigating or examining any tire. First noted should be the brand name and style name, such as Michelin XM+S 244 and the size of the tire, such as P 195 75 / R 14. Also of importance is the Department of Transportation (DOT) number, which will usually begin with the letters DOT. This is often on the serial side (inner side) of the tire. The DOT number will be similar to the following example:

DOT	HM	L9
Department of transportation	Manufacturer and plant code Michelin Tire Company Stoke-on-Trent, England	Tire size Code

ABCD	499
Group of optional symbols with the manufacturer	Date of manufacture 49th week of 1999

In this example, the two letters following the DOT number, such as HM, are symbols for the manufacturer and plant code. By looking this code up in a reference source, such as *Who Makes it and Where*, the specific plant in this example can be identified as the Michelin Tire Company, located in Stoke-on-Trent, England. The next letter and number are the tire size code. The following four letters, for example, ABCD, are optional and are manufacturers' symbols. The last three numerical digits are important, as they indicate the week and year in which the tire was manufactured. For example, the numbers 499 in this example would indicate the tire was made in the 49th week (49) of 1999 (9). The last '9' could also indicate the tire was made in 1989 or 1979.

Tires that are retreads have a slightly different DOT number on them. The retread DOT number will begin with the letters DOTR of perhaps just R. The original DOT number that was on the new tire may be removed, or may be left undisturbed, so it is possible a tire will have both a DOT and a DOTR number on it. The DOTR number consists of three letters and three numbers and would be similar to the following example:

DOTR	YPY	129
Indicates retread tire	Goodyear Retread Frederickton, NB, Canada	12th week of 1999

In this example, the three letters are a code to identify the retreading facility. A publication, *Who Retreads Tires*, lists approximately 5000 retreading facilities by their three letter code. The three numbers identify the week and year of retreading.

Tire Construction

There are three basic types of tire construction: bias, bias-belted and radial. A bias tire will have its plies running at a bias angle across the tire. A bias-belted tire will have the same plies, but will have the addition of a belt beneath the tread area. A radial tire will have the plies running in a radial direction, from bead to bead. Bias and bias-belted tires are less effective at reducing tire squirm, i.e. the degree of contraction of the tire tread while under load. Thus, an impression left by a bias or bias-belted tire may appear to have a more narrow appearance, and the grooves in the impression may appear more closed, than when looking at the static tire. Radial tires, because of the radial direction of their plies, reduce the amount of squirm, and result in almost no difference in dimension. In an examination, test impressions of the tire, made while on a vehicle, are necessary to provide a comparable standard for comparison with a crime scene impression.

Tires are built of many components, including a liner, sidewall components, the bead, plies, belts and tread rubber. These are assembled on a rotating and collapsible drum. After assembly of the components, the 'green tire', whose tread rubber contains no tread design at this point, is transferred to a mold. In the mold, under extreme heat and pressure, the tire components will be vulcanized and bonded together. A steam-pressured bladder will inflate inside the tire and press the green tire against the surfaces of the mold. It is here that the tread design and sidewall designs will be molded into the rubber.

Tire Designs and Databases

There are thousands of designs of tires, including those made for passenger vehicles, light, medium and heavy trucks, agricultural vehicles, off-the-road vehicles and motorcycles. Since the 1960s, the *Tread Design Guide* has offered a listing and photographs of most tire designs. An international version of the

Tread Design Guide is now available. This guide is published in hard copy form, but also in a more comprehensive version on a CD-ROM, which includes over 11 000 tire designs. It is subdivided by the type of tire category, i.e. ribbed tire, metric, lugged tire, etc., allowing for easier search on a PC computer format. These sources provide the investigator a method of linking a crime scene impression to a specific tire design and manufacturer. Once done, additional information of relevance can be sought from the respective manufacturer.

Some databases have also been created which contain various wheelbase, track width and turning radius data. These would enable entry of that data, as obtained from a crime scene, for the purposes of identifying potential vehicle types and brands.

Original Equipment Tires versus Replacement Tires

Original Equipment (OE) tires are those which are put on a vehicle when it is manufactured. Replacement tires are those which are purchased to replace the OE tires, or other replacement tires, when they wear out. A particular vehicle with its four OE tires will not be a rare occurrence, as many thousands of those vehicles would have been sold with the same design and size OE equipment. On the other hand, a vehicle with one or more replacement tires on it constitutes a much less frequent occurrence, as other vehicles of that type and brand are less likely to have the same brand and style replacement tires. A vehicle with three or four different designs of replacement tires in a particular size constitutes an occurrence that would probably not be repeated, owing to the large number of available designs and sizes.

Tire Size Designations

Tire sizes have been designated in a number of ways throughout the years. The tire size designations that have been used are shown below.

Older designations

Numeric: 6.45–14
6.45 = approximate section width in inches
14 = rim diameter

Alphanumeric: E R 78–14
E = load / size relationship
R = radial tire
78 = aspect ratio
14 = rim diameter

Newer designations

Metric: 195R14
195 = approximate section width in millimeters
R = radial tire
14 = rim diameter

P-Metric: P 195/75 R 14
P = passenger tire
195 = approximate section width in millimeters
75 = aspect ratio
R = radial tire
14 = rim diameter

Aspect Ratio

The aspect ratio, otherwise known as the tire profile, is the relation of the height of a tire to its width. Appearancewise, a tire with a low profile or aspect ratio, such as 50, appears to look flatter and is proportionally wider, whereas a tire with a higher profile, such as 70, appears more conventional. Today, an increasing number of passenger cars are equipped with low profile tires to increase traction and performance.

Noise Treatment

As a tire goes through the stress cycle of contraction and expansion, the vibration of the individual design elements creates harmonics, or noise. Manufacturers attempt to treat and reduce this 'noise' by varying the pitch, or size, of the design elements as they are positioned around the tire. The sounds emitted by varied pitches creates better harmonics than the sounds emitted by the same pitch repetitively. This is referred to as 'noise treatment'. Noise treatment may be as simple as creating three sizes of design elements, i.e. s, m and l (or 1, 2 and 3). In cases such as these, the elements may follow a 1,2,3,1,2,3,1,2,3, etc. arrangement around the tire. Or, they may have an arrangement of those same three sizes, but in a more random order, such as 1,3,2,2,3,1,2,3,1,1,2,3,3,2,2,1 and so forth. In other cases, more complex and varied arrangements, including more than just three different sizes, may involve a pattern such as 1,2,3,4,4,3,2,1, 3,3,2,2,2,2,3,3,4,3,2,1,1,2,3,4, and so forth. An example of such an arrangement is illustrated in **Fig. 2**. In this example, the noise treatment consists of 64 design elements of four different sizes. They are arranged in four sequences, represented by models A, B, C and D. The six tread wear indicators are also included in Fig. 2. The lower part of the figure represents a crime scene impression. Note that only one possible area of the tire could have produced the noise treat-

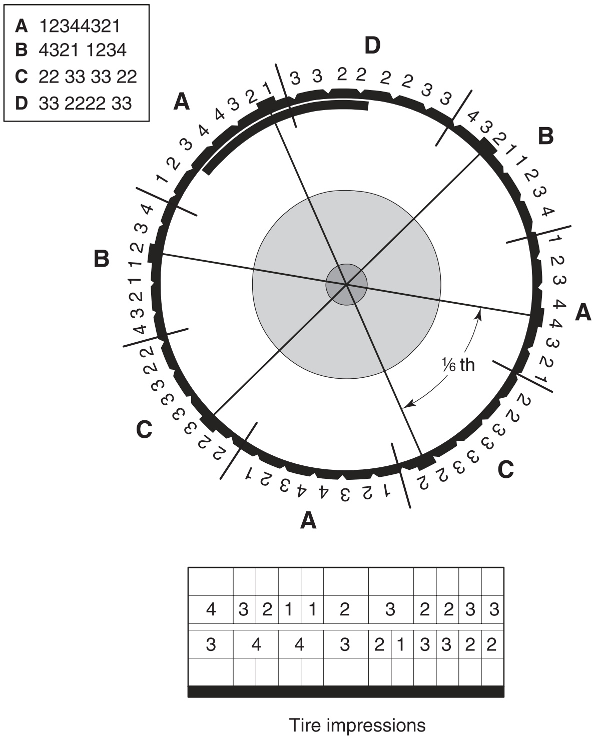


Figure 2 Noise treatment example.

ment in that impression. This is because the portion of noise treatment for model A plus the portion of model D, as shown in Fig. 2, do not repeat elsewhere in the tire.

In a *nondirectional* noise treatment, the pitch sequence will be repeated on both sides of the tire. For instance, the following illustration shows how a tire, divided circumferentially at its center line, will have the same pitch sequence on each side. No matter which direction the tire travels in, it will leave an identical repeating sequence. Examination of a crime scene impression made by a tire with nondirectional noise treatment would not enable determination of the direction of the vehicle based on this alone.

1 2 3 1 2 3 1 2 3 1 2 3 1 2 3
← direction ———Tire center line——— direction →
3 2 1 3 2 1 3 2 1 3 2 1 3 2 1

A *directional* noise treatment has a pitch sequence that lays down a different sequence of noise treatment in each direction, relative to its direction of travel. An example of this is illustrated below. An impression made by a tire with directional noise treatment would enable a determination of the direction of the vehicle from the impression, as long as the position of the tire (left or right side of vehicle) and the way the tire was mounted on the wheel (which side is out, which side is in) was known. Directional noise treatment should not be confused with a directional tire, which must be mounted in a prescribed way and whose direction can be visually distinguished.

1 2 3 1 2 3 1 2 3 1 2 3 1 2 3
← Direction ———Tire center line——— ← Direction
1 2 3 1 2 3 1 2 3 1 2 3 1 2 3

The noise treatment of tires is used during the examination of tire impressions, to assist in locating the possible position, or positions, of the tire that could have made a particular impression. An average passenger or light truck tire may have 2 m or more of tread in a full circumference impression. A crime scene photograph or a cast of a tire impression may only be 0.25–0.5 m in length. That crime scene impression represents only a small portion of the tire that made it. If the impression is short and the noise treatment is repetitive around the tire, there may be several locations on the tire that match the noise treatment in the questioned impression. Each will have to be examined closely for further evidence of wear and acquired characteristics. Recovering longer segments of impressions at crime scenes allows for a more specific determination of the portion(s) of the tire that made the impression.

Every examiner who conducts forensic tire impression examinations should be knowledgeable about tread design noise treatment and how it is significant in the comparison process.

Tread Wear Indicators

Tread wear indicators, also known as wear bars, are required in tire manufacturing. They must appear at least six times around the circumference of all tyres. They consist of rubber bars which are raised 2/32nds of an inch (0.16 cm) above the base of the grooves. In this way, as the tire wears down, the wear indicators will appear as bald strips across the remaining tread design. Tread wear indicators are visible in a two-dimensional impression if the tire is worn down to that level. They are also visible in three-dimensional impressions. Tread wear indicators, because they occur six times around a tire, appear in different portions among repeating noise treatments. For instance, a tire with a noise treatment that repeats four times will have six tread wear indicators in it. Each tread wear indicator will therefore be in a different portion of the noise treatment arrangement (Fig. 2). This can be of further assistance in locating the precise portion of the tire that made the impression.

Known Standards of Tires

Known standards of tires fall into two different categories. They include those obtained for tires that differ in design from the questioned impression, and can therefore easily be eliminated, and those that are similar in design and must be exhaustively compared with the impression with regard to the characteristics of tread design, dimension, wear and individual characteristics.

For those that fall into the category of being of different design to the questioned impressions, elimination standards need only consist of a photograph or adhesive lift of a short segment of the design. In addition, the sidewall brand name and style and size should be noted, as well as what vehicle the tire is associated with.

If the tire is similar in design to the questioned impression, the actual tire is required. When is it seized from the vehicle, its position on the vehicle (left front, right rear, etc.) and the side of the tire facing out (white wall side or black wall side) should be noted and should be marked directly on the tire *before* it is removed from the vehicle. To assist in the comparison process, full circumference known impressions of the tire must be obtained to assist in that comparison process. This is accomplished with inks or powders on long pieces of solid core chart board, or on clear polyester film placed over the chart board. This may be done on the actual suspect vehicle, or on a vehicle which will accommodate the tire.

Cases involving large trucks or trailer rigs should utilize those trucks or trailers when obtaining the known impressions. Tires should not be removed from any trucks that have double tires mounted on each side if they are to be compared with double tire impressions from the scene of the crime. Only after adequate known impressions are taken should they be removed for a more detailed examination. This is because the relative position of the tires to one another, as they have been mounted, is highly significant. Removal of the tires would lose that important aspect of the evidence.

Examination of Tire Impressions

Forensic examination first evaluates the general class characteristics of the tread design and dimension. If the tread design of the questioned impression is different from a particular tire, that tire can be eliminated. If the tread design is similar, a more detailed examination continues and requires full circumference impressions of that tire. With the full circumference impressions, the specific tread design and its dimensions, including agreement or disagreement of the noise treatment, can be compared. This examination may include superimposition of the known impression over the questioned impressions. This assists in locating the specific design, size and noise treatment features, as well as characteristics of wear and individual characteristics. If the design, size and noise treatment correspond, further comparison of the tire or tires proceeds. Any wear characteristics, such as worn away tread pattern, sipes, irregular tire wear and tread wear indicators, will assist further in locating the possible area or areas of the tire.

Finally, any random individual characteristics, including scratches, tears, cuts, stone holds and sipe damage, that appear on the tires and are clearly evident in the impression are examined. Their presence on the tire is random in nature and results in a tire that differs from others of the same tread dimension and design. The presence of random individual characteristics will contribute toward, or be the basis for, identification of that tire as making the crime scene impression. Tires are very durable, but they do quickly acquire substantial wear and damage, making them highly individual. When sufficient detail of these features is retained in the questioned impression, positive identification is readily possible.

When making examinations between a crime scene tire impression and a suspect tire, a range of conclusions is possible. It can be concluded that a tire positively made an impression. This would occur when the crime scene impression and suspect tire correspond in design, noise treatment, dimension, condition of wear and also share some random individual features. In less conclusive examinations, it is still often possible to associate a crime scene tire impression as sharing significant features in common with the suspect tire, such as design, noise treatment, dimension and condition of wear. Because of the large number of possible designs and sizes and conditions of wear, particularly when dealing with replacement tires, this category of conclusion is still very significant evidence. Finally, substantive differences in class characteristics of size or design, in the noise treatment or significant differences in the condition of wear can contribute toward, or be the basis for, elimination of the tire.

Absence of random individual characteristics, either in the crime scene impression or tire, does not constitute a basis for nonidentification, for several reasons. The impression may not have retained sufficient detail to allow for examination of those areas, or sufficient additional mileage on the vehicle may have worn off old characteristics or it may have acquired new characteristics.

Vehicle Dimensions and Turning

Tread design width

Tread design width, also known as ‘arc’ width, is the distance between the edges of the tire tread. This is sometimes difficult to measure, owing to uneven wearing of the tire, the incompleteness of the impression and other factors.

Track width (stance)

Track width, also known as ‘tire stance’, is the distance measured perpendicularly from the center of one wheel (or impression) to the opposite wheel (or impression). The track width of the rear wheels is usually different from that of the front wheels. The measurement is made more easily if the vehicle is traveling in a straight line. The front wheel measurement will change dramatically if the vehicle is turning and is therefore less reliable. The rear track, when turning, will still record accurately; however, the measurement must be made carefully, perpendicularly. If the substrate is uneven, or the impression does not record sufficiently, it is more difficult to measure either track width accurately. If a wheel of a different dimension is put on the vehicle, or if the wheel mounting changes, as in the case of mounting wheels in a reversed position, the track width for that vehicle will be changed.

The track width on trucks that have two tires mounted on each side is measured from the point that is the direct center between the two wheels on one side to the center of the two wheels on the opposite side. Should trucks of the same type, model and year be configured with different axles or different wheels, the track width will be changed. Track width and wheelbase measurements are illustrated in **Fig. 3**.

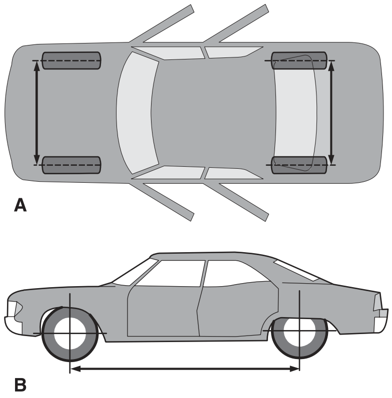


Figure 3(A) Track width: the dimension measured between the tire center lines at the ground. The front and rear track widths are not normally the same. **(B)** Wheelbase: the dimension measured longitudinally between front and rear wheel center lines.

Wheelbase

The wheelbase of a vehicle is the dimension measured longitudinally between the front and the rear wheel center lines. An equivalent to the wheelbase can be measured as the distance between the leading edges of the front and rear tire tracks, as in the case where the tires have sunk in the ground or have turned. Measurements should be taken from each side of the impression, as the leading edge of a tire, when turning, will cause this measurement to increase or decrease, depending on which side of the tire is used for the measurement.

Turning positions

Tires positioned on the rear of a vehicle, when the vehicle is turning, track to the inside of those tires positioned on the front. This is useful in reconstructing the position of tires at a crime scene and later linking similar positions to the respective tires. This is illustrated in **Fig. 4**.

Turning diameter

Every vehicle has a turning radius or turning diameter which represents the smallest radius that that particular vehicle type can turn in a circle. A crime scene impression with front wheel tracks that turn sharply can be measured and used to include or eliminate vehicles which can or cannot turn in that diameter or less. The measurement of the track should be made at its outer margin. The following formula and **Fig. 5** illustrate this procedure.

$$\text{Turning diameter} = (B^2 \div A) + A$$

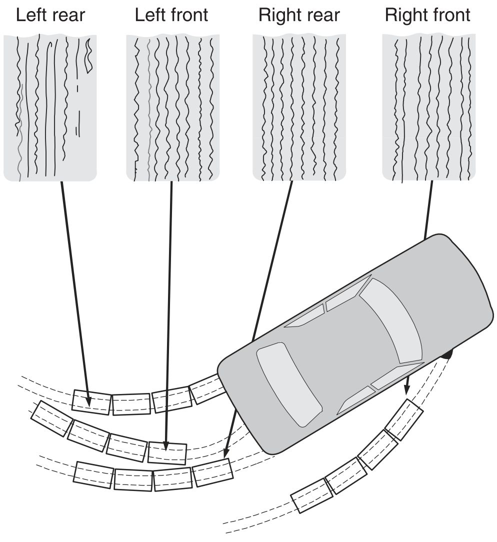


Figure 4 Turning positions. When the vehicle is turning, the rear tires track to the inside of the front tires.

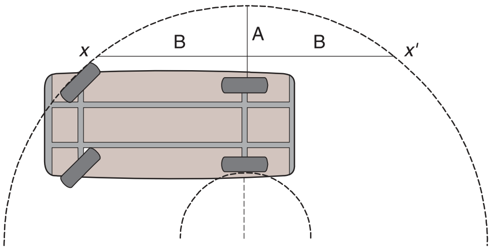


Figure 5 Turning diameter. Measurements are taken from the outer edge of the arc made by the outermost front tire. See text for details.

For example, in Fig. 5, if line x to x' equals 9 m, then, B equals 4.5 m, and if A equals 3.1 m, then:

$$\begin{aligned}\text{Turning diameter} &= (4.5^2 \div 3.1) + 3.1\\ &= (20.25 \div 3.1) + 3.1\\ &= 6.53 + 3.1\\ &= 9.63\text{m}\end{aligned}$$

Recovery Methods Specific to Tires

Photography

The general crime scene and examination quality photography of tire impressions at the scene of a crime is carried out in the same manner as for footwear marks, with one exception. That exception involves the need sometimes to document a long tire impression which is longer than that typically captured in a single examination quality photograph. This normally applies to impressions that are in excess of 0.5 m in length. Impressions longer than 0.5 m require a sequence of overlapping photographs, which upon processing and enlarging can be combined to recreate the long tire impression for comparison. To take sequenced photographs, a tape measure should be laid alongside the full length of the long impression. This measure, once laid down and once photography begins, should not be disturbed. It will not be used as a scale but only to help reconstruct the sequence and splice the photographs together. A second scale should be used and positioned in each individual tire impression segment photographed. That scale should be placed alongside, but on the same plane (level) as the bottom of, the tire impression. The photograph of each section should be approximately 0.5 m and overlap the previous photography by 6–10 cm. With this method, a long impression can be photographed in several segments, each accurately representing the respective section. The natural size enlargements of the segments can be assembled together to recreate the full impression. This method is illustrated in **Fig. 6**.

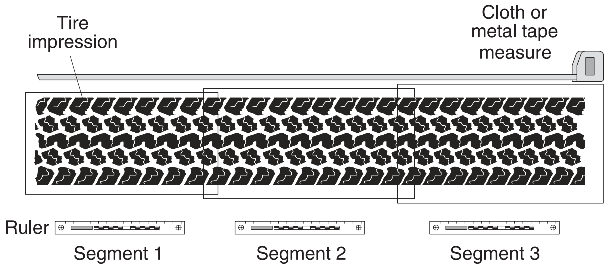


Figure 6 Documenting a long tire impression photographically.

Casting

Casts of three-dimensional tire impressions are invaluable for examination and should always be taken. They capture the contours and uneven qualities of the tire impressions, which are not always apparent or able to be recovered in photographs. In deeper impressions, sidewall treatments are often recorded, revealing potentially important information. For this reason, the casting material should be allowed to completely fill and overflow out of the tire impression. Because of the noise treatment, it is also desirable to cast the largest tire segment as is reasonably possible. Impressions that are 1 m or less usually pose no problem and should always be cast in their entirety. Larger casts can also be made, but as they increase in length, particularly if they are deep, their size and weight will rapidly become unmanageable. In cases involving long segments of one or more tire tracks from a vehicle, at least one cast, a minimum of 1 m long if possible, should be made of the best portion of each separate track. Additional casts, particularly of the more detailed segments of the tracks, should be considered, if possible. On the back of the cast, information about each impression and its relative direction should be noted. In addition, general scene photographs should be taken of the casts, once they are all poured but before they are lifted from the scene. This will also document the position and direction of the casts. Tracks made by trucks with two tires on each side, mounted side by side, should be cast as one two-tire unit. This will assure that a record of the exact relative position of those tires to each other is made.

Retread Tires

Retread tires account for a very large percentage of all tires made and sold. They are predominantly used on medium and large trucks, where the monetary savings between new tires and retread tires is significant. Some retread designs for truck tires closely simulate the original design.

Retread tires are made using one of two processes. Each of these processes leaves characteristics that

reveal the tire is a retread and may also contribute to the individuality of the tire. They are known as the *mold cured* process and the *precured* process. The mold cured process uses strips of raw rubber that are applied to the used tire carcass, which is then placed into a mold, where the new tread is formed. The resulting tire is similar to a new tire. The precured process uses premolded rubber containing the tread design, which is then bonded to the original tire carcass. It will contain a splice joint, and in some instances two splice joints, which reveal where the precured tread rubber was joined together. The position of this joint will be different from most other retread tires of this design, and will be very significant should it be retained in a crime scene tire impression.

Retread tires for passenger vehicles are not made in the quantities that original molded tires are made. Their occurrence on a vehicle and in a crime scene tire impression may therefore be far less frequent and thus more significant. Information to support the frequency of a particular tire can be obtained by contacting the retreading facility, via the retread code on the tire. A retread Department of Transportation code, explained earlier, should be present on the retread tire, but occasionally this is omitted.

See also: **Crime-scene Investigation and Examination**: Recording; Recovery of Human Remains; Preservation.

Further Reading

Bodziak WJ (1984) Shoe and tire impression evidence. *FBI Law Enforcement Bulletin* 53(7):2–12 (revised 1986).

Bodziak WJ (1996) Some methods for taking two-dimensional comparison standards of tires. *Journal of Forensic Identification* **46**(2):689–701.

Bodziak WJ (1999) *Footwear Impression Evidence: Detection, Recovery and Examination*, 2nd edn. Boca Raton, FL: CRC Press.

Bodziak WJ (1999) Forensic evidence. In: *Footwear and Tire Tread Evidence*, pp. 9-1–9-18 Sacramento CA: California District Attorney's Association.

Kovac FJ (1978) *Tire Technology*. Akron, OH: Goodyear Tire and Rubber Co.

McDonald P (1989) *Tire Imprint Evidence*. Boca Raton, FL: CRC Press.

Nause L (1987) The science of tire impression identification. *RCMP Gazette* **49**(1).

Nause L (undated) *Forensic Tire and Vehicle Track Identification*. Ottawa: Canadian Police College.

Who Makes it and Where. Tire Guides Inc., 1101-6, S. Rogers Circle, Boca Raton, FL 33487, USA: published annually.

Who Retreads Tires. Tire Guides Inc., 1101-6, S. Rogers Circle, Boca Raton, FL 33487, USA: Tire Guides Inc., published annually.

Tread Design Guide. Tire Guides Inc, 1101-6, S. Rogers Circle, Boca Raton, FL 33487, USA: published annually.

Tread Design Guide on CD-ROM. Pearl Communications, 2788 Loker Avenue West, Carlsbad, CA 92008, USA.

Payment Cards *see* **Forgery and Fraud**: Payment Cards.

PCR *see* **Deoxyribonucleic Acid**: Polymerase Chain Reaction.

PCR-STR *see* **Deoxyribonucleic Acid**: Polymerase Chain Reaction-Short Tandem Repeats.

PHARMACOLOGY

I R Tebbett, Center for Environmental and Human Toxicology, University of Florida, Gainesville, Florida, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0597

Introduction

Pharmacology is the study of the effect that a drug has on biological systems. Since 'forensic' refers to the application of science to legal issues, 'forensic pharmacology' is therefore the study of drugs and their

reveal the tire is a retread and may also contribute to the individuality of the tire. They are known as the *mold cured* process and the *precured* process. The mold cured process uses strips of raw rubber that are applied to the used tire carcass, which is then placed into a mold, where the new tread is formed. The resulting tire is similar to a new tire. The precured process uses premolded rubber containing the tread design, which is then bonded to the original tire carcass. It will contain a splice joint, and in some instances two splice joints, which reveal where the precured tread rubber was joined together. The position of this joint will be different from most other retread tires of this design, and will be very significant should it be retained in a crime scene tire impression.

Retread tires for passenger vehicles are not made in the quantities that original molded tires are made. Their occurrence on a vehicle and in a crime scene tire impression may therefore be far less frequent and thus more significant. Information to support the frequency of a particular tire can be obtained by contacting the retreading facility, via the retread code on the tire. A retread Department of Transportation code, explained earlier, should be present on the retread tire, but occasionally this is omitted.

See also: **Crime-scene Investigation and Examination**: Recording; Recovery of Human Remains; Preservation.

Further Reading

Bodziak WJ (1984) Shoe and tire impression evidence. *FBI Law Enforcement Bulletin* 53(7):2–12 (revised 1986).

Bodziak WJ (1996) Some methods for taking two-dimensional comparison standards of tires. *Journal of Forensic Identification* **46**(2):689–701.

Bodziak WJ (1999) *Footwear Impression Evidence: Detection, Recovery and Examination*, 2nd edn. Boca Raton, FL: CRC Press.

Bodziak WJ (1999) Forensic evidence. In: *Footwear and Tire Tread Evidence*, pp. 9-1–9-18 Sacramento CA: California District Attorney's Association.

Kovac FJ (1978) *Tire Technology*. Akron, OH: Goodyear Tire and Rubber Co.

McDonald P (1989) *Tire Imprint Evidence*. Boca Raton, FL: CRC Press.

Nause L (1987) The science of tire impression identification. *RCMP Gazette* **49**(1).

Nause L (undated) *Forensic Tire and Vehicle Track Identification*. Ottawa: Canadian Police College.

Who Makes it and Where. Tire Guides Inc., 1101-6, S. Rogers Circle, Boca Raton, FL 33487, USA: published annually.

Who Retreads Tires. Tire Guides Inc., 1101-6, S. Rogers Circle, Boca Raton, FL 33487, USA: Tire Guides Inc., published annually.

Tread Design Guide. Tire Guides Inc, 1101-6, S. Rogers Circle, Boca Raton, FL 33487, USA: published annually.

Tread Design Guide on CD-ROM. Pearl Communications, 2788 Loker Avenue West, Carlsbad, CA 92008, USA.

Payment Cards *see* **Forgery and Fraud**: Payment Cards.

PCR *see* **Deoxyribonucleic Acid**: Polymerase Chain Reaction.

PCR-STR *see* **Deoxyribonucleic Acid**: Polymerase Chain Reaction-Short Tandem Repeats.

PHARMACOLOGY

I R Tebbett, Center for Environmental and Human Toxicology, University of Florida, Gainesville, Florida, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0597

Introduction

Pharmacology is the study of the effect that a drug has on biological systems. Since 'forensic' refers to the application of science to legal issues, 'forensic pharmacology' is therefore the study of drugs and their

effects as they relate to the law. Examples of the types of case that may involve a forensic pharmacologist would include those involving the effect of medications and/or illicit substances on the mental and physical state of a victim or perpetrator of a violent crime; driving under the influence of alcohol or drugs; malicious or accidental poisoning; adverse drug reactions and interactions; and competency to stand trial.

The pharmacologic effect of a drug is associated with both the rate and concentration at which it reaches its site of action. These parameters in turn depend upon the absorption, distribution and elimination of the drug by the body. Before we can assess the effect that a drug has on a particular individual, we must therefore consider the relationship between dose, route of administration, metabolism and the resulting concentration of the drug in the body. Pharmacokinetics is the mathematical discipline that encompasses these processes and relates the dose of the drug administered to blood concentration, and therefore pharmacological effect.

Pharmacokinetics

Absorption

In order to produce an effect, the drug must enter the bloodstream and be carried to its site of action. The drug must therefore be absorbed across the gastrointestinal (GI) tract if given orally, from an injection site (unless given intravenously), across the skin (creams and ointments) or via the lungs (inhalations). The rate at which a drug enters the bloodstream will determine its onset of action. For example, intravenous administration results in an immediate effect because the drug is placed directly into the blood, whereas oral administration results in a gradual increase in blood concentration while the drug is absorbed across the GI-tract into the bloodstream (**Fig. 1**).

Drug absorption from the gastrointestinal tract The majority of drugs are administered orally and are absorbed via the GI tract. There are a number of factors that can influence the rate and extent of

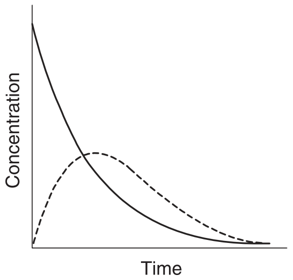


Figure 1 Plasma concentration/time curves for a drug administered intravenously (——) and orally (- - -).

appearance of intact drug into the systemic circulation after oral administration of a pharmaceutical. The steps involved in the release and absorption of a drug from a tablet are shown below. The rate and extent of appearance of intact drug in the systemic circulation depend on a series of processes. The slowest step in this series (rate-limiting step) will control the overall rate and therefore the onset of pharmacological effect. Rate-limiting steps vary for different drugs and include the rate of release of the drug from the particular dosage form, the solubility of the drug, its absorption across membranes, the rate of gastric emptying and the rate of metabolism.

Structure of the gastrointestinal tract The GI tract consists of three major regions: the stomach, the small intestine and the large intestine. As a drug passes through these regions, it experiences changes in pH, enzymes, electrolytes and surface features, all of which can influence drug absorption. The stomach mucosa contains many folds, which increase the surface area available for absorption. Although the stomach does not function primarily as an absorptive organ, its high blood supply and the fact that a drug can potentially reside in the stomach for several hours provide conditions suitable for the absorption of primarily acidic drugs. The small intestine, or ileum, is the most important site for drug absorption in the GI tract. The extremely large surface area of the small intestine results from the existence of folds in the intestinal mucosa, villi and microvilli. Most drugs reach the systemic circulation via the bloodstream of the capillary network in the villi. However, it is possible that the absorption of highly lipid-soluble drugs may occur via fat absorption pathways. The colon or large intestine, like the stomach, lacks the villi and microvilli of the small intestine, but the large intestine serves as a site for the absorption of a drug that has not been completely absorbed in the small intestine.

Mechanisms of drug transport across the gastrointestinal barrier The epithelium lining the GI tract is considered to constitute the main cellular barrier to the absorption of drugs from the GI tract. Submicroscopic, aqueous channels or pores penetrate the lipid membrane. Water-soluble substances of small molecular size, such as urea, are absorbed by simple diffusion through these channels. The majority of drugs cross the membrane by passive diffusion, whereby the physicochemical properties of the drug, the nature of the membrane and the concentration gradient determine the rate of drug transport across the membrane. The process of passive diffusion involves the partition of the drug between the aqueous intestinal contents and the membrane. Drugs entering

the bloodstream will be carried away from the site of absorption by the GI blood supply and will become diluted by distribution in a large volume of blood, distribution into body tissue, metabolism and excretion, and protein-binding in the blood. Most drugs are absorbed from the GI tract by passive diffusion; however, a few lipid-insoluble compounds, such as 5-fluorouracil, are absorbed by active transport. A carrier, which may be an enzyme or other component of the cell membrane, is responsible for effecting transfer of the drug by formation of a complex. There are several active transport systems in the intestinal tract responsible for the absorption of amino acids. Drugs which resemble these compounds will also be absorbed by the same mechanism, e.g. L-dopa, which is structurally similar to tyrosine and phenylalanine, and various vitamins and sugars.

Physiological factors influencing drug absorption A number of factors affect the absorption of a drug from the GI tract. These include:

- Surface area of the absorption site: drugs are generally absorbed more readily from the small intestine because of the extremely large surface area of this structure.
- pH of the GI fluids: the acid or base environment of the GI tract will influence the degree of ionization of the drug as it passes from the stomach to the intestine. Generally, the unionized fraction of the drug is more readily absorbed by the intestinal mucosa.
- Gastric-emptying rate: most drugs are absorbed from the small intestine; hence, any reduction in the rate at which the drug leaves the stomach and enters the small intestine may inhibit or delay absorption. Gastric emptying is particularly important for drugs that are susceptible to acid or enzyme hydrolysis in the stomach. Hunger, anxiety, body position and intake of liquids can stimulate the gastric-emptying rate. Fatty foods, high-bulk diet, depression and various drugs, including alcohol, retard it.
- Intestinal motility: most drugs are designed to be absorbed after entering the small intestine, and ideally the drug should remain in this environment for as long as possible. It normally takes 3–10 h for a drug to be carried along the length of the small intestine.
- Drug stability in the GI tract: acid–base or enzymatic hydrolysis in the GI tract can reduce the bioavailability of the drug, i.e. the amount of drug available for absorption.
- Complexation of the drug with dietary components: the GI absorption of some drugs, for example tetracycline, is reduced if taken together with dairy products, owing to the formation of an insoluble complex with calcium. Various disease states such as gastric ulcers can also affect the degree of drug absorption from the intestinal tract.

Physicochemical factors affecting drug absorption In addition to physiological factors, the physical chemistry of the particular drug also determines its rate and degree of absorption from the intestine. The dissolution rate of a drug is important because it can only be absorbed from the GI tract once it has dissolved. Dissolution rate is affected by the particle size, the presence of polymorphic forms, salts and other excipients. In general, tablets take longer to dissolve than capsules or suspensions. The greater the time taken for the drug to dissolve, the slower the onset of action. The dissociation constant (pK_a) of the drug determines its extent of ionization in the stomach and intestine and therefore the extent of absorption. The octanol/water partition coefficient of the drug gives an indication of the extent of drug absorption from the GI contents to the mucosa.

Distribution

Distribution of a drug to the tissues is subject to the same processes as absorption and is affected by the route of administration, the protein-binding of the drug, the blood supply to various organs, and the rate of metabolism and excretion of the drug. An indicator of distribution is the volume of distribution, which is determined by dividing the amount of drug in the body by the plasma concentration. If the volume of distribution for a particular drug is known, then the total amount of that drug can be determined from an analysis of the plasma concentration.

Elimination

Drugs are eliminated from the body primarily by metabolism by the liver to a more polar compound, followed by excretion in the urine by the kidneys. The rate at which a drug is eliminated from the body is described by the term clearance. Clearance generally refers to the volume of plasma from which a drug is completely removed per unit time. Plasma half-life is the time taken for plasma drug concentrations to decline by 50%. Both clearance and plasma half-life vary greatly between different drugs, and even between individuals.

Drug metabolism

Metabolism is an integral part of drug elimination. As well as facilitating excretion of the drug, it may also affect the pharmacological response of a drug by

altering its potency and/or duration of action. When a polar (or ionized) water-soluble drug is absorbed in the body, it is largely excreted unchanged by the kidneys. However, the majority of drugs are lipid-soluble to some extent. Such compounds must undergo extensive metabolism, and are converted into a more polar, water-soluble form before they can be excreted in the urine. Although metabolism is generally considered to be a detoxification process, some of the metabolic products may also have pharmacological activity and may be toxic. Examples include aspirin, which is hydroxylated to salicylic acid; diazepam, which may be hydrolyzed to temazepam and oxidized to oxazepam; and amitriptyline, which is demethylated to nortriptyline.

Pathways of drug metabolism can be nonsynthetic – for example, oxidation and hydrolysis – or synthetic – for example, conjugation with glucuronic acid or sulfate. Despite the extensive range of reactions that a drug molecule may undergo, the majority are catalyzed by membrane-bound enzymes in the liver cells (hepatocytes). For example, the cytochrome P_{450} mixed-function oxidase system catalyzes oxidations and conjugation with glucuronide. Metabolism can occur in the plasma as well as organs other than the liver, such as the GI tract, kidneys and lungs.

Oxidation Metabolic oxidation reactions include *N*-demethylation, e.g. diazepam → nordiazepam (desmethyldiazepam); hydroxylation, e.g. diazepam → temazepam; *N*- or *S*-oxidation, e.g. chlorpromazine → chlorpromazine N-oxide or chlorpromazine sulfoxide.

Hydrolysis Hydrolysis is an important route of metabolism for esters and amides, e.g. aspirin → salicylic acid. This may be as a result of acid, base or water hydrolysis, or due to esterases in the plasma or liver.

Reduction reactions Aldehydes and ketones may undergo reduction to primary and secondary alcohols, e.g. prednisone → prednisolone. Nitro compounds (NO_2) are reduced to amines (NH_2).

Conjugation reactions Oxidation, reduction and hydrolysis usually produce metabolites, which have a reactive functional group. Several of the more common conjugation reactions that mask these functional groups include glucuronide or sulfate formation, acetylation and methylation. Glucuronic acid can be conjugated with a wide range of functional groups to form either an ester (as with carboxylic acids) or ether with phenols and alcohols. *N*- or *S*- glucuronides may be produced as well as *O*-glucuronides, all of which facilitate drug elimination by the kidney. Sulfate conjugates are formed with alcohols and phenols. These are salts of strong acids and are readily excreted in the urine. Acetylation under the influence of the enzyme *N*-acetyltransferase occurs frequently with primary and secondary amines, and methylation is a common feature of the metabolism of phenols.

Factors affecting metabolism

Similar drug concentrations will not produce the same pharmacological effects in all subjects. These differences may be due to one of the following factors.

Age Young children and elderly people generally have a lower metabolic capacity compared with that of subjects between these extremes of age. The enhanced sensitivity of the very young to drugs can be accounted for by the fact that the microsomal enzymes, which are responsible for metabolism (particularly conjugation), are not fully active until several months after birth. Older children (5 years) metabolize drugs at a similar rate to adults. The dose must be lower, however, to take account of the smaller volume.

In elderly patients (over 60 years) there is a decreasing capacity for drug metabolism, as a consequence of a gradual decline in physiological efficiency. In addition, the amount of protein-binding may decrease and renal excretion may be reduced. Elderly people may therefore experience higher blood levels of a drug compared with younger patients.

Disease Diseases can affect all of the processes by which a drug is absorbed, distributed and eliminated from the body. A drug may be poorly absorbed during GI disturbances. The rate at which drugs cross tissue membranes can be altered in cardiovascular disease, which may alter peripheral blood flow. Endogenous free fatty acids released into the plasma during trauma can displace drugs that are weak acids from albumin-binding sites.

Diseases that affect the liver and/or kidneys probably have the greatest effect on drug concentrations, having a direct effect on metabolism and excretion. If liver function is impaired as a result of disease or chronic drug use, blood levels will be greatly increased.

Weight The weight of a patient affects drug concentrations in the blood, as it determines the volume into which the drug is distributed. Drug metabolism tends to be faster in males than in females.

Genetic factors The genetic control of the number of receptor sites and genetic variations in the extent of

protein-binding or the rate and extent of drug metabolism can make a marked contribution to variations in drug concentrations and responses. The inability to oxidize tolbutamide, for example, has been related to an increased incidence of cardiovascular toxicity of the drug.

Diet Diet has been demonstrated to influence the metabolism of some drugs; for example, the conversion of an asthmatic patient from a high-protein to a low-protein diet will increase the half-life of theophylline. The exposure of the patient to other drugs (particularly alcohol) can have a marked effect on the metabolism of certain drugs.

Excretion of drugs

The excretion of drugs and metabolites terminates their activity and presence in the body. They may be eliminated by various routes, with the kidney playing the major role with excretion into the urine. Drugs may also be excreted in the feces, bile, lungs, sweat, saliva and breast milk. Some drugs may be reabsorbed from the renal tubule even after having been sent there for excretion. Because the rate of reabsorption is proportional to the concentration of drug in the unionized form, it is possible to modify this rate by changing the pH of the urine. For example, acidification of the urine with citric acid causes reabsorption of acidic drugs, whereas making the urine alkaline with sodium bicarbonate causes reabsorption of basic drugs.

Pharmacology

The major classes of drugs that possess abuse potential can be classified as sedative hypnotics, stimulants and hallucinogens.

Sedative hypnotics

Opium has been used as a sedative since ancient Greek times. It is a natural product obtained from the dried latex of the opium poppy *Papaver somniferum*. Crude opium is a rich source of the narcotic analgesics codeine and morphine. It has long been known that inhalation of the vapors produced by heating opium will cause sedation and a long peaceful sleep. Unfortunately, addiction to the drug develops rapidly. Opium is still occasionally found as a street drug, although presently the majority of opiate abuse is associated with heroin, a semisynthetic derivative of morphine. Heroin has similar actions to morphine, and indeed opium, but is a more powerful analgesic and is also more addictive. Illicit heroin is either injected intravenously or smoked by inhalation of vapors produced by heating a small amount of the drug on aluminum foil. The presence of adulterants, such as quinine, strychnine and various cutting agents, including sugar, starch and even household cleaner and brick dust, make intravenous administration of this and other illicit preparations extremely hazardous.

Absorption The opiates are absorbed rapidly after parenteral administration. Oral absorption of morphine is variable and results in a bioavailability of only around 20%. It is distributed primarily in the kidneys, liver, lungs and spleen but does not accumulate in the tissues. The major metabolism of morphine is by conjugation with glucuronide to form morphine 3- and 6-glucuronides. *N*-demethylation, *O*-methylation and N-oxide formation are minor metabolic pathways. The half-life of morphine in the plasma is approximately 3 h. After parenteral administration, approximately 10% of the dose is excreted as the unchanged drug and about 70% as glucuronides.

Heroin, or diacetylmorphine, is rapidly hydrolyzed in the blood to 6-monoacetylmorphine and then to morphine. The excretion of heroin in the urine is similar to that of morphine, with the majority being converted to morphine glucuronide. The detection of low concentrations of 6-monoacetylmorphine is an indication of the use of heroin as opposed to morphine. Codeine is rapidly absorbed following oral ingestion. It is then metabolized primarily to codeine-6-glucuronide, with some demethylation to morphine and, in turn, morphine glucuronides.

Effects The pharmacological effects of the opiates are all similar, exerting their main effects by action on the central nervous system. The opiates are most effective in the treatment of moderate to severe pain. Behavioral changes induced by these drugs vary and can result in euphoria or nervousness and fear. Depression of the central nervous system (CNS) can result in respiratory depression, and respiratory failure is a common cause of death in cases of opiate overdose. Pinpoint pupils and constipation are common indicators of opiate use. Tolerance to opiates develops quickly. This is not due to an increase in metabolism but to deactivation of the opiate receptors. On withdrawal, the receptors become supersensitive and the system becomes overactive. The resulting withdrawal symptoms include exaggerated yawning and shivering and profuse sweating. Severe diarrhea and vomiting, resulting in significant weight loss, can in some cases be fatal. Goose flesh and involuntary muscle twitches have resulted in the expressions ‘cold turkey’ and ‘kicking the habit’ when referring to heroin withdrawal.

Methadone, a synthetic opiate, is widely used in the treatment of opiate addiction. Methadone merely substitutes for the heroin. Its longer half-life and the fact that it is administered orally rather than intravenously mean that it has less abuse potential than heroin. The effects of opiate intoxication can be completely blocked or reversed in an emergency situation by use of an opiate antagonist such as naloxone. Administration of such a drug will, however, immediately send the patient into withdrawal.

Other sedatives

Barbiturates are the prototype drug in this class, although they are no longer widely used, having largely been replaced with the benzodiazepines. Abuse, addiction, tolerance and dependence develop with all members of this class. The historical development of sedatives has seen the introduction of one drug after another, each with its own abuse potential. Unfortunately, it is impossible to treat anxiety, depression and insomnia using current drug therapy without the risk of addiction. The nature of the symptoms of these conditions, and the fact that drugs act directly on the CNS can lead to abuse and dependence. The barbiturates are derivatives of barbituric acid; in general, those compounds possessing substituent groups that confer greater lipid solubility have the faster onset of action and shorter half-life. Barbiturates may be administered orally or intravenously. Most are extensively metabolized, with less than 5% of a dose of the more lipid-soluble compounds amobarbital and pentobarbital being excreted unchanged. Phenobarbital, which is the most polar of the common barbiturates, is excreted as 25% unchanged in the urine. Metabolism of the barbiturates typically involves hydroxylation, oxidation or dealkylation. All of these compounds have half-lives greater than 24 h, which can result in accumulation in the body with repeated doses. Even after a single dose, barbiturates and/or metabolites are excreted in the urine for several days.

Benzodiazepines The benzodiazepines have largely replaced the barbiturates for the treatment of anxiety and are available for both oral and intravenous administration. They are probably the most overprescribed group of drugs in use today. Chlordiazepoxide, oxazepam, lorazepam and alprazolam are absorbed from the GI tract relatively slowly, and peak plasma concentrations may not be reached for several hours. Diazepam is rapidly absorbed and peak concentrations may be achieved in less than 1 h. The benzodiazepines and their metabolites bind to plasma proteins. Diazepam, which possesses high lipid solubility, is 99% protein-bound. Due to the lipophilic nature of these compounds, there is a rapid uptake into brain tissues. The benzodiazepines undergo extensive metabolism by the same routes as the barbiturates. In many cases, however, metabolism of the benzodiazepines results in the formation of other biologically active compounds. Examples include the metabolism of diazepam to temazepam and oxazepam. Plasma half-lives vary greatly, depending on the individual benzodiazepine. Flurazepam, for example, has a half-life of approximately 3 h, whereas the half-life of nitrazepam is around 24 h.

The sedative hypnotics depress the activity of all excitable tissue, particularly nerve cells. In large doses, the drugs can suppress function in cardiovascular activity. Early effects include loss of attention and concentration, impaired short-term memory and lack of coordination. Sedative hypnotics also cause withdrawal symptoms when their use is discontinued. These vary in severity, with the shorter-acting drugs typically having an earlier onset and more pronounced withdrawal effect. Symptoms include anxiety, muscle weakness and shaking. This may be followed by convulsions and psychoses, which can last for several weeks unless treated. Treatment is required to prevent morbidity from adverse effects, and the sedative is usually substituted with a long-acting benzodiazepine, such as chlordiazepoxide or diazepam, with fewer side effects. The dose of this drug is then reduced gradually over time.

Side effects include CNS depression, sedation, drowsiness, hostility, irritability and disturbing dreams. Weight gain, skin rash and headaches are also commonly encountered. Benzodiazepines are considered ‘safe’ in the sense that it is difficult, if not impossible, to overdose on them; however, benzodiazepine addiction is widespread and fits of depression accompanying its use often lead to suicide attempts. Treatment consists of switching to a longer-acting benzodiazepine that causes less pronounced withdrawal effects, and gradually reducing the dose.

Various drugs act synergistically with sedative hypnotics to produce dangerous respiratory depression and cardiac failure. Ethanol, antihistamines and monoamine oxidase inhibitors can all increase the CNS depressant activity of the sedative hypnotics. In addition, the metabolism of some drugs, such as anticoagulants, digoxin and β blockers, can be enhanced as a result of the induction of microsomal enzymes by the sedatives. This effect, if not anticipated, is potentially life-threatening.

Stimulants

The most commonly abused drugs in this group are cocaine and amphetamine. Cocaine is a natural

product present in the leaves of the coca bush, which is indigenous to the Andes mountain region of South America. Like opium, cocaine has a long history of use and abuse, and is still chewed by South American Indians for its stimulant effects. Illicit cocaine may be found in two forms: the hydrochloride salt, and the free base or 'crack'. The hydrochloride is usually snorted up the nose with absorption through the mucous membranes, whereas the free base is smoked and absorbed through the lungs. The bioavailability of cocaine varies depending on its route of administration, with only 20–40% of an oral or intranasal dose being absorbed compared with nearly 100% via the lungs. Insufflation results in a euphoric high after about 30 min, compared with 75 min after an oral dose and 8 s by smoking. Cocaine is primarily metabolized by hydrolysis to benzoylecgonine, and by the action of esterases to ecgonine methyl ester. *N*-demethylation of cocaine results in the formation of norcocaine, which retains biological activity. Cocaine has a plasma half-life of around 40 min, and benzoylecgonine of about 90 min.

Cocaine acts by activation of the brain's pleasure centers, which are dependent on dopamine neurotransmission. It blocks the reuptake of dopamine from the synapse, which results in euphoria, increased motor activity and psychotic symptoms. Apart from its stimulant effects, cocaine also causes increased heart rate, cardiac arrhythmias and peripheral vasoconstriction. Consequently, sudden changes in blood pressure may result in death due to cardiac failure and cerebral hemorrhage. Cocaine addiction is treated with antidepressants, such as bromocriptine, or tricyclics, such as desipramine.

Amphetamine Amphetamines are generally used as oral preparations, resulting in a widespread distribution of the drug. Amphetamine undergoes several routes of metabolism: hydroxylation, *N*-dealkylation, oxidative deamination, *N*-oxidation and conjugation of the nitrogen. Approximately 50% of the drug is excreted unchanged in the urine. For most amphetamines, the major route of metabolism is oxidative deamination. Unlike most drugs, chronic administration of amphetamine inhibits its metabolism, resulting in an increase in the amount excreted unchanged. The half-life of amphetamine in the blood is dependent on urinary pH, and is approximately 4–8 h if the urine is acidic and about 12 h in individuals whose urinary pH is not controlled.

Amphetamine is similar in its pharmacological effects to cocaine, and is also thought to act via its interaction with dopamine. Amphetamine promotes the release of dopamine and norepinephrine from presynaptic neurons and blocks the uptake of catecholamines. In common with cocaine, amphetamine causes restlessness, stimulation, appetite suppression, paranoia and psychosis. Other stimulants with amphetamine-like activity include methamphetamine, methylenedeoxyamphetamine (MDA) and methylenedeoxymethamphetamine (MDMA). As little as 3 h after cessation of chronic amphetamine use, the patient falls asleep. Sleep may last for several days and is followed by increase in appetite, anxiety, depression, irritability and lethargy. Psychosis, including paranoia, hallucinations and suicidal tendencies, may develop, from which it may take years to fully recover.

Hallucinogens

Cannabis, or marijuana, is the most commonly abused psychoactive substance. The fruiting and flowering tops and the leaf material of the *Cannabis sativa* plant contain the highest concentrations of the active compounds, or cannabinoids. Dried plant material may be mixed with tobacco and smoked. Hashish is a resinous extract of the cannabis plant, which is similarly smoked and contains higher concentrations of cannabinoids. Cannabis has a rapid onset of action, causing sedation and disruption of space–time perception. Although smoking is the most common route of administration, it also has some activity if taken orally. Smoking results in onset of drug action within seconds, with peak tetrahydrocannabinol (THC) concentrations occurring after 7–10 min and declining over 2–3 h. THC is highly lipophillic and is stored in body fat, which may result in its being detected in the urine of chronic users for several weeks after the last dose. THC undergoes rapid metabolism to 8,11-dihydroxy-THC and an active metabolite 11-hydroxy-THC, which is further metabolized to Δ^9-THC acid and excreted in the urine. The half-life of THC in the plasma is about 20–36 h.

Cannabis causes increases in heart rate and peripheral blood flow, and bronchodilation. Other effects include increased body temperature, slower EEG activity, decreased intraocular pressure, increased appetite, drowsiness, visual distortions, diminished concentration and attention span, and decreased coordination. Chronic use is reported to result in mutagenic effects, as well as causing a reversible inhibition of the immune system. Psychotic delusions and paranoia may also occur with prolonged use. A common effect of chronic cannabis use is depression, which may be severe and result in suicide attempts. Anxiety and sexual dysfunction have also been associated with its use. 'Amotivational syndrome' is a controversial term used to describe personality

changes associated with chronic use of cannabis. It is defined as a lack of interest in work or productivity.

Lysergic acid diethylamide (LSD) LSD is a semisynthetic substance produced from ergot alkaloids. It was originally developed as a possible treatment for schizophrenia, but its side effects limited its usefulness. It is one of the most powerful mind-affecting substances known – a typical dose as low as 50 μg being capable of producing an effect that will last for several hours. LSD is well absorbed after oral dose, with psychological effects becoming apparent after 45 min and peaking at 3 h after administration. Effects may last for 8–12 h. The half-life in plasma is approximately 2 h. LSD undergoes extensive metabolism in the liver, with only about 1% being excreted unchanged. 2-Oxylysergide is a major metabolite, with 13-hydroxylysergide and 14-hydroxylysergide also being produced.

LSD is thought to produce its effect by interacting with serotonin in the brain. The action mixes up the senses, commonly known as a 'trip'. Vivid hallucinations, such as seeing music as colors, are common. Although it is thought to be impossible to overdose on LSD, deaths attributed to accidents while individuals are under the influence of the drug are relatively common. Flashbacks or the recurrence of hallucinogenic effects are also a symptom associated with LSD. They are probably due to a very small amount of the drug being absorbed by the brain tissue and later released into the system, causing a second hallucinogenic effect.

Psilocin and psilocybin Psilocin and psilocybin are naturally occurring hallucinogens present in the *Psilocybe* genus of mushrooms. They have an action similar to that of LSD but are less active. Ten to fifteen mushrooms (about 2 cm in height) will induce a hallucinogenic effect when taken orally. Psilocybe mushrooms can also be dried and smoked with tobacco or the drug may be extracted with boiling water to produce an infusion. Hallucinogenic effects generally become apparent within 30 min of ingestion. Psilocybin is readily absorbed from the GI tract, and is metabolized by dephosphorylation to the hallucinogen psilocin. The duration of hallucinogenic effect is several hours after a single dose.

Phencyclidine (phenylcyclohexylpiperidine, PCP) Phencyclidine was originally marketed as a veterinary anesthetic. It is used illicitly as a white powder, which is taken orally, smoked together with marijuana or inhaled in combination with cocaine. It may also be injected intravenously. PCP's onset of action accordingly varies from a few seconds with intravenous use and smoking, to 20 min after oral ingestion. The half-life is variable and is reported to range from 10 to 96 h, with effects reported as lasting between a few hours to days. Like amphetamine, the rate of elimination depends on urinary pH. Metabolism occurs primarily in the liver, with glucuronidation and hydroxylation being the principal pathways. Approximately 50% of the administered dose is excreted unchanged.

Symptoms appear rapidly and resemble schizophrenia. Users are a risk to themselves and to others when under the influence of this drug. Euphoria, depression, agitation, violence, hallucinations, paranoia, panic and suicidal tendencies are common effects. PCP users also experience increased strength and a decreased sense of pain, making them extremely dangerous and difficult to control. Eventually they lose consciousness and go into a coma. Hypertension and tachycardia are also associated with the use of PCP, which may result in cardiac failure. PCP-induced psychosis can last for up to 4 weeks, during which time the addict has to be kept sedated because of violent behavior. Phencyclidine is thought to act by increasing glucose utilization in the brain. This is probably initiated via a receptor and may involve acetylcholine. PCP intoxication is characterized by reddening of the skin, enlarged pupils, delusions, amnesia, nystagmus, excitement, arrhythmias, paranoid psychosis and violent behavior. With high doses, convulsions lead to death.

Solvent abuse

Teenagers who are unable to obtain other drugs or alcohol frequently abuse solvents. Most solvents cause CNS depression, sedation, intoxication, euphoria and hallucinations. The solvent is placed in a plastic bag and sniffed, or a cloth soaked in the solvent is placed over the nose and mouth and inhaled. Toluene is the main solvent that is abused. It causes rapid CNS depression that mimics the effects of narcotics such as morphine. It also causes liver damage and renal and cardiac failure, which may lead to convulsions and death. Solvent intoxication is initially similar to the effects of alcohol, with excitation, euphoria, nausea and vomiting. Further inhalation results in progressive depression of the CNS, accompanied by disorientation and confusion. With continued exposure, ataxia, stupor, seizure and cardiorespiratory arrest follow.

Sources of solvents include adhesives (benzene, toluene, xylene, hexane and heptane), aerosol sprays (ethanol, isopropanol, toluene and xylene), antifreeze (glycols, methanol and isopropanol), nail polish (acetone, acetates and alcohol) and glue (petroleum

distillates, ethyl acetate and toluene). The inhalation of most solvents results in effects similar to those described for toluene. In addition, methanol results in abdominal cramps, headache and muscular weakness. Methanol is metabolized to formaldehyde, which causes damage to the optic nerve. Isopropanol initially causes irritation of the eyes and nose but can be fatal in sufficient doses. Butyl and amyl nitrate cause ischemia and hypotension and may result in convulsions and cardiovascular collapse. Inhalation of gasoline results in euphoric effects in less than 5 min, but symptoms may persist for up to 6 h. Cardiac arrhythmias, hallucinations and neurological problems are typical. Numerous cases of sudden death have been associated with solvent abuse. This is largely due to CNS depression followed by respiratory collapse.

See also: **Drugs of Abuse**: Body Fluids; Classification, including Commercial Drugs; Ante-mortem Blood; Urine; Hair; Analysis, **Education**: An International Perspective.

Further Reading

Gerald MC (1974) *Pharmacology: an Introduction to Drugs*. Englewood Cliffs, NJ: Prentice Hall.
Giannini AJ and Slaby AE (1989) *Drugs of Abuse*. Oradell, NJ: Medical Economics Books.
Goodman LS and Gilman A (1996) *The Pharmacological Basis of Therapeutics*, 9th edn. New York: Macmillan.
Tilstone WJ and Stead AH (1986) Pharmacokinetics: metabolism and interpretation of results. In: Moffat AC (ed.) *Clarke's Isolation and Identification of Drugs*, pp. 276–305. London: Pharmaceutical Press.

PHONETICS

H Hollien, IASCP, University of Florida, Gainesville, FL 32611, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0792

Introduction

There is little question but that the forensic sciences are dynamic and that their dramatic growth parallels current advances in technology and system operation. DNA analysis and the computerization of fingerprints are but two examples. Forensic phonetics also is currently developing. It is defined as a professional specialty based upon the utilization of knowledge about human communicative processes and the application of the specialized techniques which result. As with the forensic sciences in general, the focus here is on law enforcement and the courts.

Forensic phonetics consists of electroacoustical analysis of speech signals that have been transmitted or stored, and the analysis of the communicative acts themselves. The first involves the proper transmission and storage of spoken exchanges, the authentication of tape recordings, the enhancement of speech on tape recordings, speech decoding and the like. The second area involves issues such as speaker identification and the process of analyzing speaking behaviors to obtain information relative to the emotional, physiological or psychological states of a talker. **Figure 1**

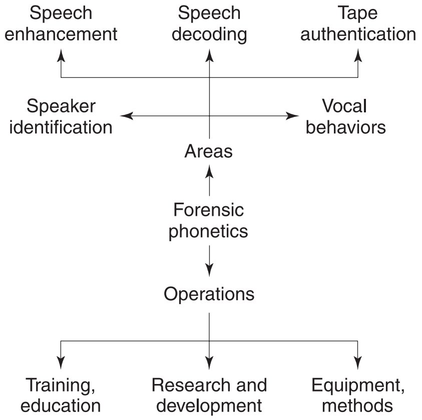


Figure 1 The nature and scope of forensic phonetics. The five content areas most basic to the specialty are listed in the upper third of the figure; the three major activities at the bottom. Not shown are any of the secondary or interface areas discussed in the text.

provides a visualization of the field. Note that the signal versus human analysis dichotomy has been dropped and five basic areas have been identified. Also listed are the specialized operations that lead to progress in the cited fields. Not included are several secondary areas; in those cases, forensic phonetics is usually interfaced with forensic linguistics (for language analysis) or audioengineering (for nonhuman signal analysis). They were not included in the figure, as they will not be discussed in this brief review. What is covered here will be problems with: (1) the integrity of captured utterances (and related), (2) the accuracy/completeness of messages, (3) the identification of the human producing the utterances, and (4) determination of the various physical states experienced by a person. These areas will be defined and the professional response discussed.

Authentication of Tape Recordings

The problem of tape and/or video recording integrity is a rather serious one as a challenge (at any time) relative to their validity suggests that someone has tampered with evidence or has falsified information. Yet tape recordings are quite vulnerable and changes to them can be either intentional or accidental. The purpose of authentication, then, is to determine if the target recordings can be considered valid and reliable or whether they distort what actually happened.

But, what is authenticity? To be authentic a tape recording must include the complete set of events that occurred and nothing must have been added, changed or deleted during the recording period or subsequently.

For purposes of this discussion, the focus will be on assessing analog tape recordings. Analysis of digital or video recordings is also based upon these procedures; however, specialized tests must be added. Evaluations are divided into two parts: the physical examination and signal assessment.

Physical examination

Five tasks are carried out prior to initiation of the physical examination:

1. A log should be structured.
2. A high-quality copy should be made.
3. The tape should be listened to for familiarization purposes.
4. The evidence pouch should be examined for openings.
5. The tape housing should be checked for identification marks.

The first assessment is to determine if the correct amount of tape is on the reel. Usually the manufacturer's estimates are a little casual (closeness to the reel edge, the number of 'minutes' the reel or cassette contains). While small variations are of little importance, large ones suggest manipulation. The second examination is focused on the cassettes or reel. The housing is examined for pry marks, damage to the screw heads, etc. Negative 'evidence' here is cause for concern. Third, the tape must be examined for splices – either adhesive or heat. If any are found, it must be assumed that editing has taken place and the area around the splice must be carefully examined. Fourth, the recorder on which the tape purportedly was made must be examined and tested. The procedures here involve making a series of test recordings under conditions of quiet, noise and speech, and repeating them while serially operating all switches associated with the system. Later, evaluations will determine if the tape is actually the original. For example, the electronic signatures found on the tape must be the same as those produced by the equipment (that is, if the tape is to be judged authentic). They also permit determination of stop-and-start activity within the tape, over-recordings and so on. Finally, some of these issues can be assessed by additional methods. An example is where the 'tracks' made by the tape recorder drive mechanism are examined; another involves the magnetic recording patterns themselves (as seen by application of powders or solutions).

Laboratory examination

Once copies of the 'originals' are observed to be good duplicates, they may be taken to the laboratory for further tests. First, the recording is listened to critically and all questionable events are logged. This process is repeated 5–10 times. Questionable events include clicks, pulses, thumps, clangs, rings, crackles, etc. They also may involve loss of signal, abrupt shifts in ambient noise, shifts in amplitude, inappropriate noises, and so on. Apparent breaks in, or difficulty with, context, word boundaries or coarticulation must also be identified and recorded. Indeed, a tape recording must be rigorously examined for *any* event that could suggest it has been modified.

Once all of the questionable events have been located and logged, it is necessary to systematically identify and explain each of them. Most will be innocuous – for example, a telephone hang-up or a door closing; dropouts could result from change in microphone position or operation of interface equipment. On the other hand, any of these 'events' could signal a malicious modification.

The procedures by which the questionable events

can be explained vary from careful listening to complex signal analysis. Any number of procedures can be applied: time-amplitude, wave and/or spectral analyses are examples. In **Fig. 2** an abrupt change in a waterfall-type spectrum signals that the tape has probably had a section removed. Even noise processing and flutter analysis can be useful. The goal is to identify and/or explain every questionable event that can be found on the tape. It is a long and tortuous process and, while most of the suspicious events will be found to be 'innocent', it is important not to miss evidence of manipulation. Finally, it must be remembered that even if it is demonstrated that the tape recording has been altered, the reason for such modification cannot be specified. It could have been either accidental or intentional; the examiner has no way of knowing.

Special analyses may be required; video tapes may have to be authenticated; the tape may be digital (DAT) or may have been 'minimized' (presumably to prevent the invasion of privacy). These problems are met by special techniques.

Finally, there are three reasons for testing the integrity of a tape recording: (1) to challenge its authenticity; (2) to defend it; or (3) simply to determine what happened. However, it must be stressed that the evaluations must be totally independent of the needs or desires of the agency that requests them. Indeed, all examinations must be conducted thoroughly and ethically.

Difficult Tapes and Transcripts

Tape recordings used for legal and law enforcement purposes are rarely of studio quality; indeed, they often are of severely limited fidelity. The two main sources of difficulty here are noise and distortion. In turn can be caused by the use of poor quality equipment, poorly trained operators and/or hostile environments. Equipment inadequacies include: limited frequency bandwidth, harmonic distortion, internal noise, amplitude limits, and so on. The most common sources are inadequate pickup transducers (microphones, 'bugs', telephones), poor quality recording equipment, very slow recording speeds, very thin tapes and/or inadequate transmission links. The second factor, operator error, is self-explanatory.

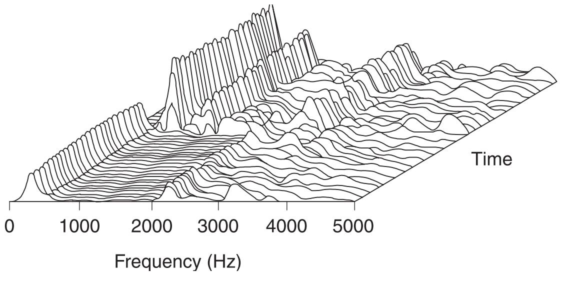


Figure 2 A waterfall plot of a rerecorded splice. Note the very abrupt break in the serial spectra.

Third, masked or degraded speech can result from environmental factors; they include: external 'hum', wind, automobile movement, fans/blowers, clothing friction. In other instances, speech is blocked by other talkers, music or 'forensic' noise. Whatever the cause or causes, the events must be analyzed and compensatory actions taken; that is, if reasonably good transcripts are to accrue.

Enhancing speech

Basically, the two remedies are speech enhancement and decoding, but they are preceded by the log, tape duplication and listening to the tape for familiarization purposes.

The first process involves filtering. If there is a substantial amount of noise at either (or both) the extreme lows and highs, speech may be enhanced by subjecting the tape recording to a band-pass filter with an appropriate frequency range (i.e. 300–3500 Hz). Second, if a spectrum analysis shows that a noise source is producing a relatively narrow band of high energy at a specific frequency, a notch filter may be employed. Third, comb filters can help when noise exists within the frequency range for speech. They can be used to modify continuously the spectrum of a signal by selectively attenuating relatively narrow frequency bands throughout its range. Of course, not very much filtering can be carried out here as speech elements can be removed along with the unwanted sounds. Finally, it must be stressed that these procedures are best carried out with digital filters or with systems that have been developed to compensate for multiple noise problems (**Fig. 3**). In any event, digital filters (either hardware or software) are advantageous, as distortion is reduced to a minimum and rebound cannot occur.

Since binaural listening is helpful, the equipment should be organized to permit stereo. Further, variable-speed tape recorders also can be useful. There are two kinds of systems here; each is used to produce a different effect. With the first type, a manual increase or decrease of tape recorder speed is possible. These units are especially useful in situations where the speed of the original recording was varied for some reason. The second type of recorder has compensatory circuits that allow the perceived speech to appear undistorted. Finally, there are yet other techniques that can be useful. Filling short dropouts with thermal noise is one. Yet others (sophisticated techniques such as linear predictive coefficients, deconvolution, bandwidth compression, etc.) can be helpful.

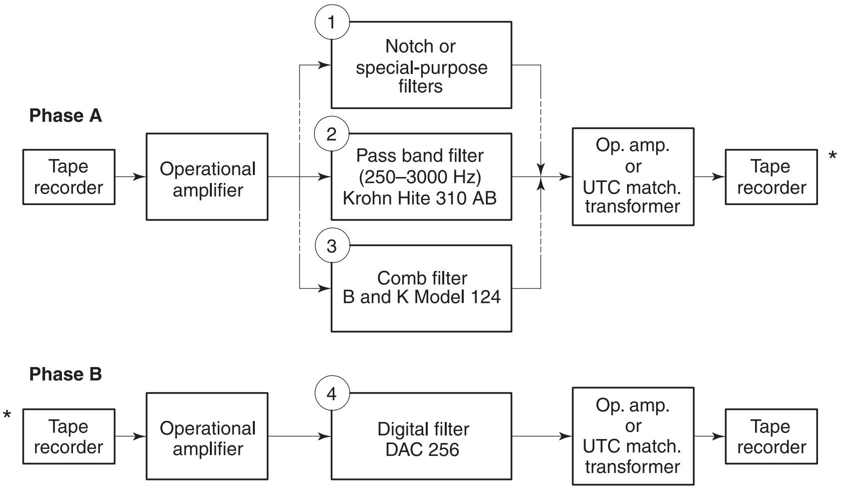


Figure 3 A block diagram demonstrating how analog and/or digital filtering can be accomplished. The three analog filters (1–3) can be used singly or cascaded; however, they would not be employed in parallel (as shown). A typical configuration would be to place a comb filter and a digital filter (both isolated) in a series. *For single phase use, remove these tape recorders and one operational amplifier.

However, at present, they are of limited value in the ordinary forensic situation as they are quite complex, time-consuming and expensive.

Speech decoding and transcripts

Speech enhancement, of course, is only the preliminary step in the development of optimum transcripts. Very often specialists must be employed to make them. These individuals are professionals who can deal effectively with factors such as voice disguise, dialects (and/or foreign languages), very fast speech, multiple speakers, stress, psychosis, fear, fatigue, intoxication, drugs and health states. These, and other problems, can make the decoding task quite demanding. But what approaches are useful here?

As would be expected, intense listening procedures are required, as is the use of the best tapes possible and equipment with good fidelity (including binaural headphones). Initial concentration on relatively easy-to-decode portions of the tape can aid in developing perceptual cues for problem areas. For example, it is well known that words have an effect upon one another, that context aids decoding and that determination of one of the words in a two-word series often can aid in the correct identification of the other.

Decoding can sometimes be aided by the use of graphic displays of the speech signal. Particularly useful to this approach are sound spectrograms of the time–frequency–amplitude class. For example, displays of this type can be useful in estimating the vowel used (from the formants), the consonant used from among several possible, and the consonant spoken by comparison of its energy distributions to identifiable productions. As would be expected, a variety of techniques are available to assist the specialist with the decoding process.

Finally, basic knowledge about, and skills in applying, many of the phonetic and linguistic realities relative to language and speech are also quite helpful. For example, a basic understanding of the nature and structure of vowels, consonants and other speech elements is necessary. So too is a good understanding of word boundaries, word structure, dialect, linguistic stress patterns and paralinguistic elements of speech (fundamental frequency, vocal intensity, phoneme/word duration, and so on). Especially useful is a thorough understanding of coarticulation and its effects on words and phrases. In short, speech/language systems can be differently employed for purposes of decoding.

Decoding mechanics

Just as the approaches to the decoding process must be highly organized, so too must the reporting structure. That is, it is not good enough to just write down the words and phrases heard (and verified); it also is necessary to explain what went on during the entire recording period and what is missing. Codes are useful here but they must be used consistently and systemically. A couple of examples should suffice. If sections are found where the speech is audible but not intelligible, more information must be provided than simply ‘inaudible’. Rather, it might look like this ‘(10 seconds of inaudible speech by talker A)’ or ‘(6–8 words of unintelligible response)’. Another requisite is to include all the events: ‘Talker A:

"Help me." (footsteps, door closing, two gunshots, loud thump)'.

Once the decoding process has been completed and confusions resolved, the transcript can be structured in final form and carefully evaluated for errors. As many prosecutors and defense attorneys have discovered (sometimes to their distress), it is only those transcripts that are accurate and reliable that can be effectively utilized in the courtroom.

Speaker Identification

Speaker identification should not be confused with speaker verification. The second of these two processes (verification) results when a known and cooperative person wishes to be recognized by means of an analysis of a sample of his or her speech. In this case, exemplars have already been made available and 'reference sets' have been developed for the talker. Speaker verification techniques are most useful to industry, government and the military; they are only occasionally of use to law enforcement agencies. Moreover, since speaker identification is, by far, the more difficult task of the two, any technique that will work for identification purposes should work even better for verification. In any case, speaker identification is where attempts are made to identify an individual from his or her speech when he or she is unknown and when anyone within a relatively open-ended population could have been the talker. Considerable research has been carried out on this issue; it has focused on three areas: (1) aural/perceptual approaches, (2) 'voiceprints' and (3) machine/computer approaches. Only the first and third of these will be reviewed, as 'voiceprints' have been shown to be inadequate to the task (and in the extreme) and have been pretty much discarded (even in the United States). Hence, it would not appear useful to outline an invalid technique in this very short review.

At least primitive efforts in voice recognition probably antedate recorded history; they have continued down through the millennia. But what is happening these days? Courtroom testimony is common but most of it is by some type of listener. Some Courts permit witnesses to testify (about the identify of a speaker) only if they are able to satisfy the jurist that they 'really know' that person. Many courts permit qualified specialists to render opinions. Here, a sample of the unknown talker's speech (usually an evidence tape) must be available, as must an exemplar tape recording of the suspect. The professional carries out an examination and then decides if the two tapes contain, or do not contain, the voice of a single person. Finally, a third type, here, involves earwitness line-ups or 'voice parades'. They will be considered first.

Earwitness identification

An earwitness line-up (or voice parade) is ordinarily conducted by the police. They ask the witness to listen to the suspect's exemplar incorporated in a group of 5–6 recorded speech samples obtained from other people. Ordinarily, the witness is required to listen to all the samples and make a choice as to which one is the perpetrator. However, this approach has come under fire and research is currently being carried out to assess its validity. Moreover, guidelines for auditory line-ups are being developed. Basically, an earwitness line-up is defined as a process where a person who has heard, but not seen, a perpetrator attempts to pick his or her voice from a field of voices. As a procedure, it must be conducted in a scrupulously fair manner. Detailed criteria relative to the procedures involved are now available.

Aural–perceptual approaches

As would be expected, a great deal of research has been carried out in an attempt to discover the nature and dimensions of speaker identification when conducted by listeners. A substantial number of the relationships here are now reasonably well understood. Briefly, correct speaker identification is enhanced if:

- the listener knows the speaker.
- listeners' familiarity with the talker's speech is reinforced occasionally.
- high-fidelity samples are available.
- the speech samples are large and varied.
- noise masking is minimal.
- those listeners who are naturally quite good at the task are used.
- listeners use the 'natural' characteristics found within speech/voice.
- examiners are phoneticians crosstrained in forensics.

Naturally, negatives here will tend to degrade identification accuracy.

Use of specialists It is clear that forensic phoneticians are able to assist attorneys and law enforcement personnel with speaker identification tasks. Indeed, there now are a number of them who can demonstrate valid analysis techniques, superior identification rates and extensive graduate training in the phonetic sciences and especially in forensic phonetics.

Many phoneticians attempt to determine identity by assessing speech segmentals (i.e. speech sounds). Such procedures are usually successful if the practitioner establishes a robust set of criteria to follow, is reasonably well trained and is experienced in this type

of analysis. But an even more powerful technique is one that is rigorously structured and uses both suprasegmentals (assessment of voice, prosody, frequency patterns, vocal intensity, quality and so on) plus the segmentals. This approach is illustrated in **Fig. 4**.

The method cited involves obtaining multiple speech samples of the unknown (evidence tape) and known (exemplar) speakers and placing them in pairs on an evaluation tape. The pairs are played repeatedly and comparisons are made of one speech parameter at a time (Fig. 4). Each characteristic (pitch patterns, for example) is evaluated continually until the judgment is made. The next parameter is then assessed and the process replicated until all possible comparisons have been completed. At that time, an overall judgment is made and a confidence level generated. Ordinarily, the entire process is independently repeated several times over a period of several days.

If the overall range of scores falls between 0 and 3, a match cannot be made and the samples were probably produced by two different people. If the scores fall between 7 and 10, a reasonably robust match is made. Scores between 4 and 6 are generally neutral but are actually a little on the positive side. Incidentally, if foils are used in the process, the confidence level is substantially enhanced, that is, if the mean scores polarize (i.e. 0–3 or 7–10).

Reasonable amounts of data now are available about people's ability to make aural–perceptual speaker identifications. Even though it has some limitations, fairly good results can be expected if the task is highly structured and if the auditors are well trained professionals who are able to demonstrate good competency.

Score sheet 0 = U– K least likely; 10 = U– K most likely

		Score	Range
1. Pitch			
(a) Level	0 . . . . 5 10		
(b) Variability	0 . . . . 5 10		
(c) Patterns	0 . . . . 5 10		
2. Voice quality			
(a) General	0 . . . . 5 10		
(b) Vocal fry	0 . . . . 5 10		
(c) Other	0 . . . . 5 10		
3. Intensity			
(a) Variability	0 . . . . 5 10		
4. Dialect			
(a) Regional	0 . . . . 5 10		
(b) Foreign	0 . . . . 5 10		
(c) Idiolect	0 . . . . 5 10		
5. Articulation			
(a) Vowels	0 . . . . 5 10		
(b) Consonants	0 . . . . 5 10		
(c) Misarticulations	0 . . . . 5 10		
(d) Nasality	0 . . . . 5 10		
(e)	0 . . . . 5 10		
6. Prosody			
(a) Rate	0 . . . . 5 10		
(b) Speech bursts	0 . . . . 5 10		
(c)	0 . . . . 5 10		
7. Other			
(a) Speech disorders	0 . . . . 5 10		
(b)	0 . . . . 5 10		
Mean			

Figure 4 A form useful in the recording of scores resulting from aural–perceptual speaker identification evaluations. Each parameter is rated on a continuum ranging from 0 (definitely two individuals) to 10 (the samples are produced by a single person). The resulting scores are summed and converted to a percentage which, in turn, can be viewed as a confidence level estimate U–K, unknown–Known.

Machine/computer approaches

The speaker recognition issue changes radically when efforts are made to apply modern technology to the problem. Indeed, with the seeming limitless power of electronic hardware and computers, it would appear that solutions are but a step away. Yet such may not be the case. While the problem is simple, attempts to solve it are complex and messy. Most of the effort here has gone into speaker verification (described above), which is neat, clean and much simpler. None the less, it is necessary to address the identification problem.

A good way to develop a machine-based system is to establish a structure and test it. For example, a group of vectors or relationships could be chosen and then researched, as shown in **Fig. 5**. The process proceeds step-by-step to assessments of multiple vectors, distorted speech, field assessment, etc. While there may be no (single) vector which is robust enough to permit efficient identification under any and all conditions, a profile approach usually is effective. A number of potentially useful programs in the speaker identification area have been initiated; however, even though some were promising, virtually none of them was sustained. An exception is the SAUSI (semiautomatic speaker identification system) program being carried out at the University of Florida.

The first step was to identify and evaluate a number of speech parameters which might make effective identification cues. It was discovered early that traditional approaches to signal processing appeared lacking; hence, natural speech features were adopted. This decision was supported by results from early experiments, the aural–perceptual literature and the realization that people routinely process heard sounds using these very features. They included speaking fundamental frequency (SFF), voice quality (LTS), vowel quality (VFT) and temporal patterning or prosody (TED); several others were tried but proved lacking.

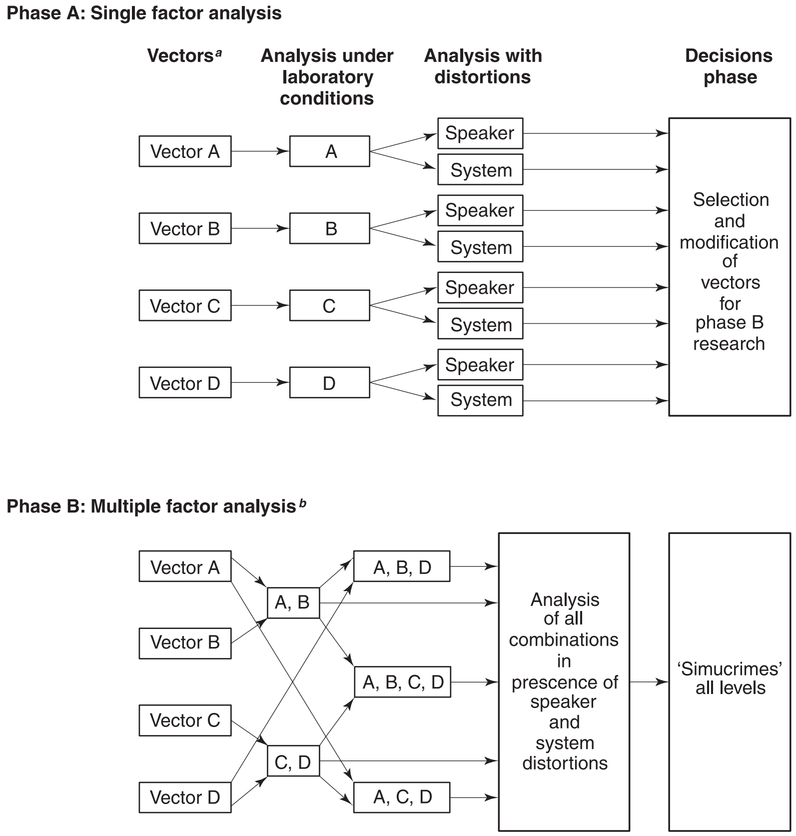


Figure 5 A structured research approach for the development of computer-based speaker identification systems. [a] Plus others; [b] illustrative only as each vector will be studied in all possible combinations.

A second perspective also emerged: it was noted that no single vector seemed able to provide high levels of correct identification for all the degraded speech encountered. Hence, those being studied were normalized, combined into a single unit and then organized as a two-dimensional continuum or profile. This rather harsh procedure addressed the severe limitations imposed upon the identification task by the forensic model (i.e. one referent; one test sample within a field of competing samples), as the research design forced matches (or nonmatches) from a fairly large collection of voices (6–25 in number). After the profile is generated, the process is replicated several times. The final continuum usually consists of the data from 3–5 rotations and includes a summation of all vectors. Hence, any decision made about identity is based on several million individual comparisons (factors, parameters, vectors, rotations).

Figure 6 is a practical illustration of how this procedure works. The evaluation here (a real-life case) involved two unknowns (X,Y), two knowns (K,N) and a number of foils. The unknown X was selected as himself (as necessary for validity), with the data for unknown Y placing him very close. The knowns and foils were in a different part of the continuum. Additional rotations compared both unknown Y and known K to the rest. What is suggested here is that the two unknowns were probably the same speaker but that none of the knowns or foils were either X or Y. Case outcome provided (nonscientific) support for this judgment.

A summary of the more recent experiments which support this approach is given in **Table 1**. The first of

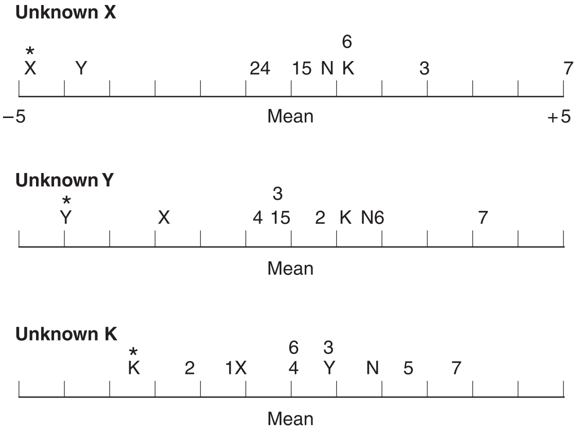


Figure 6 An example taken from a real-life investigation but one where the unknowns (X and Y) may or may not be the same person and where K and N are the suspects. Note that, in each of the three rotations, the target talker (starred) was found to best match himself.

Table 1 Summary data from three major experiments evaluating the SAUSI vectors under three conditions

Conditions and study	*Vectors*[a]				
	TED	LTS	*SFF*	*VFT*	*Sum*
High fidelity					
Hollien (1988)	62	88	68	85	90
Hollien *et al*. (1993)	63	100	100	100	100
Jiang (1995)	82	100	80	100	100
Noise					
Hollien *et al*. (1993)	64	90	77	92	100
Jiang (1995)	76	94	76	96	100
Telephone passband					
Hollien *et al*. (1993)	55	92	90	88	98
Jiang (1995)	58	100	90	96	100

TED, temporal patterning or prosody; LTS, voice quality; SFF, speaking fundamental frequency; VFT, vowel quality.
[a] Values are percent correct identification of 25 or more adult male talkers where all samples were subjected to three complete rotations.

these (1988) involved a large number of subjects but only laboratory samples (high fidelity). Note that none of the individual vectors provided 100% correct identification. The second part of this project (not shown) was designed to test the proposition that SAUSI would eliminate a known speaker if he was not also the unknown; it did so and at a level of 100% correct elimination. The second set of experiments (1993) involved a large number of subjects, plus separate replications for high fidelity, noise and telephone passband. Here, the upgrading of the vectors resulted in marked improvement for all conditions. Finally, the SAUSI vectors were further upgraded and replications of the 1993 experiments were run. As can be seen, correct identifications were strikingly higher for all conditions and the correct identification level reached 100% for all summed rotations. In short, it appears that machine-based approaches to speaker identification are feasible.

Vocal Behaviors

A number of psychological or physiological states can be deduced (sometimes) by speech and voice analysis. Of these, two have been selected for brief review. They are psychological stress and intoxication.

While determining how a person is feeling just from hearing his or her voice is not something which is very easy to do, there are times when there is little else to go on; hence, this area is quite important.

Stress

First, it should be noted that the term 'stress' denotes a negative psychological state, but is it fear, anger or anxiety? A reasonable definition of stress would appear to be that it is a 'psychological state which results as a response to a perceived threat and is accompanied by the specific emotions of fear and anxiety'.

It has long been accepted that listeners can identify some emotions (including stress) from speech samples alone, and do so very well. If this is true, what are some of the vocal correlations of this psychological state? First, increases in pitch or speaking fundamental frequency appear to correlate with stress increments. However, if this relationship is to be functional, the subject's baseline data should ordinarily be available as it (the behavior) actually results as a shift from the norm. Second, while frequency variability is often cited as a correlate of stress, it actually is a poor predictor. Third, vocal intensity is another acoustic parameter that may correlate with psychological stress; however, the data here are a little 'mixed'. Nevertheless, the best evidence is that vocal intensity tends to increase with stress. Fourth, while identification of the prosodic speaking characteristics related to stress is a fairly complex process, the temporal pattern of fewer speech bursts appears to correlate with it. Finally, an important recent finding is that speaker nonfluencies appear to increase sharply with stress.

A predictive model of the vocal correlates of stress has been developed (**Fig. 7**). As may be seen, changes occur in speaking fundamental frequency, nonfluencies, vocal intensity, speech rate and the number of speech bursts; however, it should be remembered that information of this type will be of greatest value when it can be contrasted with reference profiles for that person's normal speech.

The psychological stress evalvator It would be a mistake to leave this area without some reference to

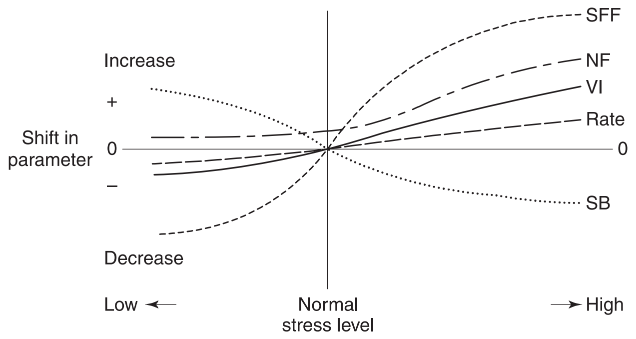


Figure 7 Model of the most common shifts in the voice and speech of a person who is experiencing psychological stress. SFF, speaking fundamental frequency; NF, nonfluencies; VI, vocal intensity; SB, number of speech bursts per unit of time.

'voice stress analyzers'. These devices, it is claimed, can be used to detect *both* stress and lying. It is without question but that the legal, law enforcement, intelligence and related agencies would benefit greatly from the availability of an effective method for the detection of stress and, especially, deception. Taking deception first: can lies be detected by any means at all? is there any such thing as a lie response? Perhaps Lykken has articulated the key concept here. He argues that, if lies are to be detected, there must be some sort of a 'lie response', a measurable physiological or psychological event that always occurs. He correctly suggests that, until a lie response has been identified and its validity and reliability have been established, no one can claim to be able to measure, detect and/or identify falsehoods on anything remotely approaching an absolute level. But has such a lie response been isolated? Simple logic can be used to test this possibility. For example, consider what would happen if it were possible to determine the beliefs and intent of politicians simply from hearing them speak. There would be no need for trials by jury as the guilt or innocence of anyone accused of a crime could be determined simply by asking them: 'Did you do it?' Consider also the impact an infallible lie detection system would have on family relationships! The answer seems clear.

Yet the voice analysts claim they can detect falsehoods and do it with certainty. They market a number of 'systems' for that purpose. How do these devices work? Unfortunately, it is almost impossible to answer this question as the claims are quite vague. One explanation is that the systems utilize the microtremors of a human's muscles. Such microtremors do exist in the long muscles of the body; however, there is very little chance that they either exist in (or can affect) the antagonistic actions of the numerous and complexly interacting respiratory, laryngeal and vocal tract muscles. Indeed, there is substantial evidence that they do not. And is the presence of stress equivalent to lying in the first place? A myriad of such questions can be asked but, at present, there are virtually no valid data to support the claims of the voice stress evaluators; rather the great preponderance demonstrates that they are quite invalid. Indeed, it appears that the PSE is an even greater fraud than are voiceprints.

Alcohol–speech relationships

Almost anyone who is asked to do so probably will describe the speech of an inebriated talker as 'slurred', 'misarticulated' or 'confused'. But do commonly held stereotypes of this type square with the results of reality and/or research? More importantly, are there data which suggest that it is possible to determine a person's sobriety solely from analysis of his or her speech? An important question, yet only limited research has been reported.

The rationale for an intoxication–speech link is clear-cut. Since both cognitive function and sensorimotor performance can be impaired, so too can the speech act, which results from operation of a number of high-level integrated systems (sensory, cognitive, motor). Moreover, it can be argued (from research) that articulation is degraded, speech rate slowed and perception of impairment raised as intoxication increases. Degradations in morphology and/or syntax have also been reported, as have articulatory problems. Perhaps more importantly, it has been found that:

- speaking fundamental frequency level is changed and its variability increased;
- speaking rate often is slowed;
- the number and length of pauses is often increased;
- amplitude or intensity levels are sometimes reduced;
- nonfluencies are markedly increased.

The basic problem with virtually all of the relationships cited is that they are variable and the reasons for

this are not clear. Moreover, any number of other behavioral states – stress, fatigue, depression, effort, emotions and speech/voice disorders – can complicate attempts to determine intoxication level from speech analysis. And such determinations might not be possible in the first place unless the target utterances can be compared to that person's speech when sober.

Having recognized the confusions and contradictions associated with the intoxication–speech dilemma, a team at the University of Florida developed a research program focused on resolution of these conflicts. New approaches designed to induce acute alcohol intoxication were employed; here, subjects received doses of 80 proof rum or vodka mixed with both a soft drink (orange juice, cola) and Gatorade. The subjects drank at their own pace but breath concentration levels (BrAC) were measured at 10–15 min intervals. The approach was efficient, with nausea and discomfort sharply reduced, serial measurements permitted and intoxication level highly controlled. Moreover, large groups could be (and were) studied, with subjects participating in all procedures related to their experiment. Data were taken at 'windows' or intoxication levels (ascending or descending), including (among others) BrAC 0.00 (sober), BrAC 0.04–0.05 (mild), BrAC 0.08–0.09 (legal) and BrAC 0.12–0.13 (severe). Subjects were carefully selected on the basis of 27 behavioral and medical criteria. After training, they were required to produce four types of speech at each intoxication level. Included were: (1) a standard 98 word oral reading passage, (2) articulation test sentences, (3) a set of diadochokinetic gestures, and (4) extemporaneous speech. As may be deduced from these descriptions, very careful and precise procedures were carried out for all conditions and levels. Analyses included auditory processing by listeners (drunk–sober, intoxication level, etc.), acoustic analysis of the signal, and various classification/sorting (behavioral) tests.

A number of relationships have already emerged. First, it appears that auditors tend to overestimate speaker impairment for individuals who are only mildly (to moderately) intoxicated. On the other hand, they tend to underestimate the level of involvement for subjects who are severely intoxicated (**Fig. 8**). Second, it appears possible to simulate accurately rather severe levels of intoxication, and even reduce the percept of intoxication if an inebriated individual attempts to sound sober. Moreover (and surprisingly), there seem to be only minor gender differences and few-to-none for drinking level (light, moderate, heavy). Perhaps the most powerful data so far are those observed for large groups of subjects in the 'primary' experiments (**Fig. 9**). As can be seen, they show shifts for all of the speaking characteristics measured except vocal intensity. Note also that speaking fundamental frequency (heard pitch) is *raised* with increases in intoxication level – a relationship which was noted by clinicians but not by previous researchers. Perhaps the most striking relationship of all is that between nonfluencies and intoxication level. The correlation here is a very high one and the pattern seen in Fig. 9 has been confirmed. While some variability exists, the predictable relationships are that speech is slowed down as intoxication increases and the number of nonfluencies rises sharply for the same conditions.

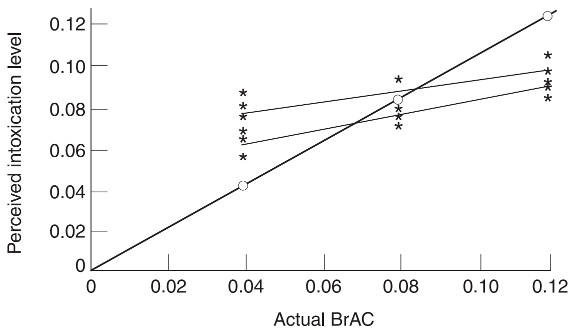


Figure 8 Perceived intoxication level contrasted to the physiologically measured levels (45° line with circles) from sober to severely intoxicated (BrAC 0.12–0.13). Four studies are combined for the top set (35 talkers, 85 listeners) and two for the lower (36 talkers, 52 listeners). Note the overreaction to the speech of people who are mildly intoxicated and underrating of those who are seriously inebriated.

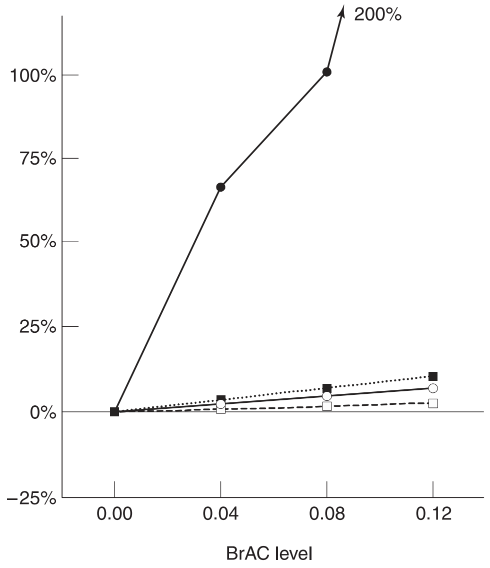


Figure 9 The shifts in several speech parameters as a function of increasing intoxication. The increase in F0 and reduction in speaking rate (increased duration) are actually statistically significant. However, they are dwarfed by the dramatic shift in nonfluencies. ○, F0(SFF); ■, duration; ●, nonfluencies; □, vocal intensity.

In Summary

The new forensic sciences subspecialty of forensic phonetics can be seen as a dynamic and growing area. Some of its elements are quite well established, while others need refinement. Of course, it must be stressed that, just as with all relatively new fields, much is to be learned about what can and cannot be accomplished by application of the methods and procedures proposed or in use. Fortunately, appropriate baseline materials have been, or are being, developed by relevant practitioners and scientists situated both in America and Europe.

Acknowledgments

The research the author describes in this section has been supported primarily by the United States National Institutes of Health, Office of Naval Research, Army Research Office and Justice Department, plus the Dreyfus Foundation and the University of Florida. Grateful thanks are extended to them.

See also: **Voice Analysis**.

Further Reading

Arbuckle T, Chaikelson J and Gold D (1994) Social drinking and cognitive function revisited. *Journal of Studies on Alcohol* **55**:352–361.

Atal BS (1974) Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification. *Journal of the Acoustical Society of America* **55**:1304–1312.

Bakis R and Dixon NR (1982) Toward speaker-independent recognition-by-synthesis. *IEEE Proceedings ICASSP*. pp. 566–569.

Bull R and Clifford BR (1984) Earwitness voice recognition accuracy. In: Wells GL and Loftus E (eds) *Eyewitness Testimony: Psychological Perspectives*. Cambridge: Cambridge University Press.

Chin SB and Pisoni D (1997) *Alcohol and Speech*. San Diego: Academic Press.

Dean JD (1980) The work of the home office tape laboratory. *Police Research Bulletin* **35**:25–27.

DeJong G (1998) *Earwitness Characteristics and Speaker Identification Accuracy*. PhD thesis, University of Florida.

Doherty ET (1976) An evaluation of selected acoustic parameters for use in speaker identification. *Journal of Phonetics* **4**:321–326.

Edie J and Sebestyen GS (1972) *Voice Identification General Criteria*. Report RADC-TDR-62-278. Griffiss AFB, NY: Rome Air Development Center, Air Force Systems Command.

Hicks JW Jr and Hollien H (1981) The reflection of stress in voice – 1: Understanding the basic correlates. *Proceedings of the Carnahan Conference on Crime Countermeasures*, pp. 189–194. Lexington, KY.

Hindermarch I, Kerr JS and Sherwood N (1991) The effects of alcohol and other drugs on psychomotor performance and cognitive function. *Alcohol and Alcoholism* **26**:71–79.

Hollien H (1990) *Acoustics of Crime*. New York: Plenum Press.

Hollien H (1992) Noisy tape recordings in Forensics. *ESCA Proceedings: Speech Processing in Adverse Conditions*. pp. 167–170. Nice, France:

Hollien H (1993) *Forensic phonetics*. In: Wecht's *The Forensic Sciences* New York: Matthew Bender.

Hollien H (1995) The future of speaker identification: a model. *Proceedings of the 13th International Congress Phonetic Sciences*: vol. 3, pp. 138–145. Stockholm.

Hollien H (1996) Consideration of guidelines for earwitness lineups. *Forensic Linguistics* **3**:14–23.

Hollien H and Hollien PA (1995) Improving aural–perceptual speaker identification techniques. In: Braun A and Koester J-P (eds) *Studies in Forensic Phonetics*, pp. 87–97. Trier: Wissenschaftlicher Verlag.

Hollien H and Jiang M (1998) The challenge of effective speaker identification. *RLA2C Avignon 1998*, 2–9.

Hollien H and Majewski W (1977) Speaker identification by long-term spectra under normal and distorted speech conditions. *Journal of the Acoustical Society of America* **62**:975–980.

Hollien H and Martin CA (1996) Conducting research on the effects of intoxication on speech. *Forensic Linguistics* **3**:107–127.

Hollien H, Hicks JW Jr and Oliver LH (1984) A semiautomatic system for speaker identification. In: Borowski V and A Koester J P (eds) *Neue Tendenz Amgerwandten, Phonetik III*, pp. 89–106. Hamburg: Helmut Buske.

Hollien H, Geisson L and Hicks JW Jr (1987) Voice stress evaluators and lie detection. *Journal of Forensic Sciences* **32**:405–418.

Hollien H, DeJong G and Martin CA (1998) Production of intoxication states by actors: perception by lay listeners. *Journal of Forensic Sciences* **43**:1163–1172.

Jiang M (1995) *Experiments on a Speaker Identification System*. PhD thesis. University of Florida.

Jiang M (1996) Fundamental frequency vector for a speaker identification system. *Forensic Linquistics* **1**:45–58.

Kuenzel HJ (1987) *Spechererkennung*. Heidelberg: Kriminalistik.

Kunzel H (1994) On the problem of speaker identification by victims and witnesses. *Forensic Linguistics* **1**:45–58.

Lykken D (1981) *A Tremor in the Blood*. New York: McGraw-Hill.

Nolan JF (1983) *The Phonetic Basis of Speaker Recognition*. Cambridge: Cambridge University Press.

Scherer KR (1986) Voice, stress and emotion. In: Appley H and Trumbull R (eds) *Dynamics of Stress: Physiological, Psychological and Social Perspectives*, pp. 157–179. New York: Plenum Press.

Shipp T and Izdebski K (1981) Current evidence for the existence of laryngeal macrotremor and microtremor. *Journal of Forensic Sciences* **26**:501–505.

Stevens KN (1971) Sources of inter- and intra-speaker variability in the acoustic properties of speech sounds. *Proceedings of the 7th International Congress on Phonetic Sciences*, Montreal, pp. 206–232.

Wallgren H and Barry H (1970) *Actions of Alcohol*, vols 1 and 2. Amsterdam: Elsevier.

Yarmey AD (1986) Verbal, visual and voice identification of a rape suspect under different levels of illumination. *Journal of Applied Psychology* **71**:363–370.

PHOTOGRAPHY AND DIGITAL IMAGING

Contents

Overview

Digital Imaging Enhancement

Overview

L Klasén, National Laboratory of Forensic Science, Linköping, Sweden

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0598

History of Images in Forensic Science

Images have always played an important role in forensic science and normally we refer to photos or digitally stored images. During the last decade, photography and digital imaging have increased in importance, as computers provide us with the tools to analyze images.

An image can be defined as a 'visually recognizable pattern'. The human visual system has the ability to interpret the pattern presented on the retina and we use what we have learned for recognizing what we see (**Fig. 1**). We are rather good at analyzing images and now we want computers to be at least as good as we are. That is not an easy task. The key components we use for dealing with visual informations are image acquisition, preprocessing, processing, analyzing, interpretation and visualization of the analysis result, storage, transmission and database search.

In essence every contact leaves a trace, which is one of the fundamental principles of forensic science. Images that have been directly captured at a crime scene can be used as evidence. In this case the trace consists of the radiated and reflected photons from the objects at the scene. These photons are passed through the optics, captured onto the sensor surface, quantified into an electric signal processed through the electronic devices, to give a signal that can be displayed or converted into a magnetic track on a videotape (**Fig. 2**).

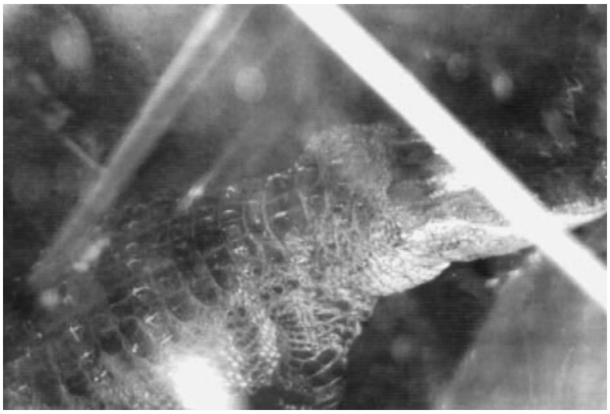


Figure 1 (see color plate 45) A poor image of a crocodile enhanced from a negative, in which we usually recognize its typical textured skin.

Figure 2 Images digitized from a surveillance system recording.

Images are widely used in other forensic applications, for instance in the examination of firearms, toolmarks, shoeprints, fingerprints and questioned documents. However this can be considered as indirectly captured traces from the crime scene. There are many specialized imaging techniques and analytical methods for the purpose, referring to their respective forensic applications.

Initially, sketches were used for illustrating a crime scene. Conventional photographic techniques were introduced in the middle of the nineteenth century and have been used ever since. Although digital imaging gives the impression of being a relatively new technique, it actually dates back more than 30 years and has passed through several phases and driving forces. One of the driving forces was the introduction of digital images as a new tool. During World War II 'electronic and digital calculating machines' appeared, and this started the era of computers. One great advantage was that these machines could be used for large two-dimensional data structures, such as images. Another driving force was the expansion of the scientific fields. New solutions led to breakthroughs and formed new scientific areas. Optical character recognition, blood cell counting and target recognition in aerial photographs are some examples.

The digital images were initially generated from other media, but during the 1970s the first digital satellite image was produced. It was quite a complex task to view, store and handle these large amounts of data. In the mid 1980s there was a breakthrough for both research and applications, which provided systems for processing and analyzing digital images and image sequences. The success was followed by a backlash, caused by unrealistic optimism and by underestimation of the complexity of the human visual system.

Then the use of image processing systems at forensic laboratories began, which provided methods to deal with surveillance photos and videos. The main purpose was to acquire and handle images and to provide printouts. Image enhancement methods were also introduced.

In the late 1980s, tools for image analysis and image-based measurement became available; these were refined during the 1990s to become more precise, complex and user-friendly. Today, any computer can be equipped with hard- and software to deal with digital images at a reasonable price and computer processing time. The research has also reached a more mature status, partly by merging several scientific fields, as image processing and computer graphics. In the late 1990s software specialized for forensic application became available. This overview of methods and applications only mirrors the state-of-the-art as new tools and techniques will continuously be introduced.

Image Acquisition Techniques

Imaging refers to the process of capturing data by imaging devices. Images can be analog, digital or hybrid. Examples of analog images are photographs, optical images and holograms. The most common TV and video standards are examples of the combination of hybrid techniques, as there are a discrete number of TV-lines but each TV-line is an analog representation of the scene.

The imaging techniques have varied over the years from being artistic, photographic and magnetic to electronic. It is therefore no surprise that evidential images come in a variety of formats.

Photographs can be considered as one important class, and digital or digitized images another.

In all image acquisition stages information is more or less lost or distorted and artifacts might be introduced.

Photographic techniques

Conventional photographic techniques are important for documentation and for developing and securing traces, by providing high quality photos. Light composition is important in the imaging stage, and by varying the intensity and direction of the light the trace can be enhanced. Photographic techniques also have the capability to extend the observations outside the visual range, by varying the wavelength of the lighting source from UV and IR regions and transform the trace into the visual range. In contrast to many digital images, the high resolution provides the ability to enlarge interesting regions from positives and negatives. Photos can be developed from the negatives with varying techniques, for example enhanced contrast, intensity or color mapping for better visibility.

Another possibility is to electrically charge the trace, such as for dust particles to visualize shoeprints.

Digital imaging techniques

A digital image consists of a discrete number of picture elements, so called pixels. Each pixel carries information from the depicted scene, digital representation of intensity and usually also color. The pixels are ordered in a rectangular grid in the image, normally referred to as rows and lines. Digital images can be acquired by scanning photos or negatives, transmitted from digital still cameras, digitized video, data stream from digital camera and video or copied data files. There are several ways to achieve and distribute data files, such as downloading from the Internet or by using other telecommunication drivers.

Most imaging techniques result in high quality images, but to analyze evidential images often means dealing with low quality images and unknown imaging devices. It is therefore essential to be aware of typical device characteristics. This can be quite a task due to the large number of ways of achieving a digital image and the variety of imaging devices, copying techniques and telecommunication channels. Some major problem areas will be discussed as the effects of poor imaging can fruitfully be used in the investigation process.

Sharpness and resolution mainly characterize the quality of an acquired image. Useful terms are, the transfer functions of the optics and imaging systems, respectively (MTF) and (OTF), used to measure how a known input signal is affected by the components of the imaging system. Another measurement is the point spread function (PSF), which provides an alternative measure of the optical quality. Other functions used are the contrast sensitive function, and the signal to noise ratio (SNR). For evidential images these parameters are mostly not controllable and they therefore remain unknown. At present there are no well-established methods for quantifying the image quality acquired from unknown systems.

The visual information is affected at several levels, such as by photon effects, blur caused by motion, lens distortion, thermal noise, jittering in video signals or distorted image ratio induced by frame grabbers. Shot noise and thermal noises are normally not the main problems, provided the recording environment is normal. When charge coupled device sensors (CCD-sensors) are used, high temperature or low illumination seriously affect the image quality.

Low spatial resolution is considered the main problem compared to the quality of photos. In some way this can be compensated for by the large amount of images generated in a video recording. Digital still cameras continuously offer improved resolution, and additional data, such as date, time of exposure and type of camera. Comments might be stored as text and in complementary audio files. Presently there is a wide range of image formats and coding schemes used for digital still cameras and digital video, some are compatible and some are not. Comparing the image size and color depth to the true size of the image file might indicate if unknown compression and restoration techniques have been used. There is also the possibility that enhancement functions available in many cameras and videos have been used.

Image processing methods used for corrections of distortions and artifacts require a great deal of caution and knowledge. When images are used for identification of persons or objects by superimposing two digital images, it is essential to investigate and if necessary correct for distortion before the analysis. Otherwise the shapes and contours might remain affected, which actually are the features used for identification. Correction techniques are used to compensate for distortions caused by the optics, such as removal of radial lens distortion. Distortion of the image aspect ratio from a digitizing board is another error that can be corrected quite easily.

Videotapes from surveillance systems are perhaps the most common source of evidential images. The quality of surveillance system recordings indicates the problems and weaknesses of recording philosophy. The images often suffer from low resolution, poor lighting conditions, blur caused by optics and motion, and physical stress on videotapes. These are currently the main problems rather than the lack of scientific analytical methods. Distorted or lost synchronization pulses cause further problems. For this purpose, time-based correctors or similar devices stabilizing the synchronization pulses are valuable. To avoid further degradation of the videotape, the first step of the analysis is normally to digitize the sequences of interest. This is important in order to keep up the quality and thereby retain the fine details.

There is a similar chain of stages from digital images to printed paper copies affecting the image information. This will not be discussed further here, but it might be important to bear in mind for the preparation of material for court presentations. This particularly holds for the processing of color and the effects of raster and printer techniques.

Methods and Applications

Analytical methods are used to extract the information that is not directly obvious to the eye. For this purpose additional information is derived from the images, such as numerical values or processed images.

The processed images or image features are used to associate an image directly captured at the crime scene, to the corresponding features of a suspect. The interpretation of the analysis result is evaluated to finally determine the association of a suspect with a crime scene. Science provides powerful tools, bearing in mind that complex solutions are complex to use and require cautious implementation, validation and interpretation of the result.

Images can be processed and analyzed either by biological systems, e.g. by the human visual system, or by technical systems. Technical image processing can be divided into analog and digital methods, which can further be either linear or nonlinear (**Table 1**). Enhancement of photos by photographic techniques is an example of analog and linear processing. Digital images in combination with science enable the use of sophisticated nonlinear processing and analytical techniques. Mathematics is the basic platform, with fundamentals as algebra, calculus, statistics, transformation and optimization theory. Other relevant scientific areas are briefly explained in a following section.

The dominant casework methods are image enhancement techniques and person or object identification (**Table 2**). Examples where image information is valuable are identification of persons involved in robberies, fraud or any other crime captured on surveillance systems, identification of pedophiles, distribution of child pornography on photo, video, or by telecommunication channels such as the Internet. There are also other useful features of the evidence to be found in the storage medium. The processing of audio signals is closely related to that of visual images. Apart from images and audio there might be electrical signals, magnetic tracks, physical and chemical features or data in text files and file headers. These features give information valuable for tracing imaging devices as well as editing, copying, production and distribution order. Images are also important to illustrate and document the analytical methods and procedures. Court presentation is one such area of importance.

Authenticity

By investigating the authenticity the aim is to determine whether the information of an image medium has been technically altered. There are many software packages and video-editing systems for commercial

Table 1 Image processing and analytical methods

Biological systems	Our human visual system (as one example)	Highly complex processing of visual information	
Technical systems	Analogue processing	Linear methods Non-linear methods	low complexity
	Digital processing	Linear methods Non-linear methods	low to high complexity processing

Table 2 Image analysis methods

Imaging	Recording devices
	Distribution of image
Image authentication	Technical alteration and manipulation
Image processing	Enhancement
	Enlargement
	Restoration
	Processing with multiframes
	Correction of moving or defocused objects
Person identification	Identification from facial images
	Craniofacial identification and facial reconstruction
Object recognition	Identification from images
Image measurement	2D measurements, 3D reconstructions
	Calibrated or uncalibrated cameras
Technical investigations	Image information
	Physical features
	Electrical signals
	Magnetic fields
	Reconstruction of damaged image media
Other methods	Image interpretation
	Course of event, simulations
	Image transmission and database search

use that provide functions for editing images and image sequences. Methods used for copyright protection can be used to examine image authenticity. Imaging devices might also use coding schemes in the acquisition process, such as JPEG (**Fig. 3**), which are useful in the investigation process. The artifacts and distortions vary with optics, imaging device, format, coding technique and noise patterns. The altering of image context often causes irregularities in distortion patterns, and these abnormalities are used to determine authenticity (**Fig. 4**).

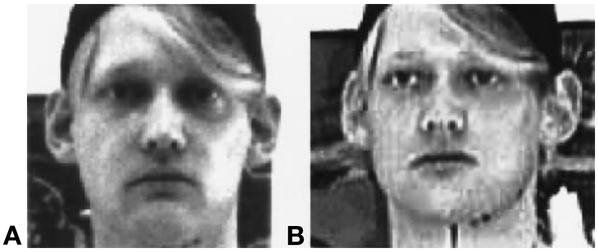


Figure 3 Artifacts from image compression caused by a JPEG coding scheme in the right image.

There is a lack of automatic methods for analyzing authenticity, which today is a mostly manual task. Other features than the image can also be used, further described in the section of technical investigation of image storage media and recording devices. To determine if a certain camera has been used to expose certain negative slides, the shape and size of the bar between the exposures on the negatives can be analyzed (**Fig. 5**), as well as the film transfer mechanism and the 'toolmarks' induced onto the negative. Recently methods for identification of digital still cameras have been presented.

Image processing

The main purpose of image processing is to produce images that are easier to interpret, visually or technically. Enhancement methods dominate and are used to extract additional information and improve the visual quality to better suit the eye. These methods can be used to deal with motion distortion or defocused objects. Automatic processing of image sequences, enlargement, restoration of lost data files or damaged image media are other methods. Enlargement of images from video is not as easy as for photographs (**Fig. 6**), due to low spatial resolution. The use of neighboring pixels and also neighboring images in a sequence has proved to be successful for enlargement as well as for enhancement, referred to as super-resolution, mosaicing and frame fusing. Data can also be recovered from erased data files or damaged image medium, but recovery of polluted or damaged videotapes is a time-consuming manual work.

Person identification

The identification of persons by anthropometry is not a new science. The anthropometric method of Alphonse Bertillion is well known, but nowadays computer-aided methods have led to a breakthrough in this area, and have extended the analysis not only to the anthropometrical measurements and features, e.g. iris pattern, face, ears, fingerprints, hands, feet, or the posture, but also to the behavior of the human body in motion. Facial identification by image superimposition, three-dimensional (3D) measurements and comparison of a 3D-reconstruction of faces or skulls to facial images is the most frequently used image-based method (**Table 3**).

Identification from facial features can be based on visual and metric comparison, computer-based 2D or 3D comparison and alignment or by automatic pattern recognition methods. When identifying a person from facial images it is essential to be aware of the functionality of the human visual system, as there are many ways to fool our brain, which for example has the capability to fill in uncompleted segments. Automatic pattern recognition is currently less useful for forensic purposes as it requires a highly controllable imaging phase, which is not the case for acquiring

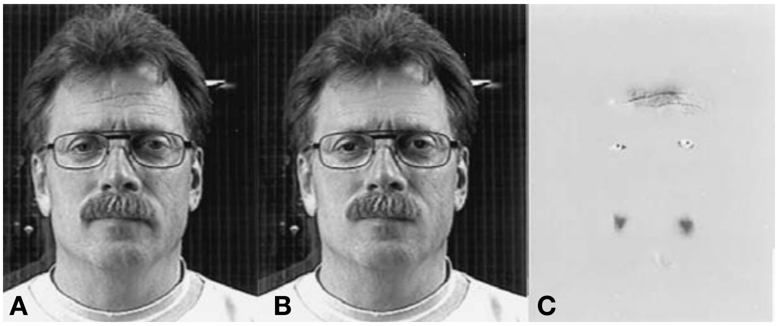


Figure 4 **B**, an altered image of **A**. In **C**, a nonlinear mapping of the grayscale has enhanced the abnormalities caused by the alteration.

A B

Figure 5 **A, B** Variations in the spacing between two exposures on a negative slide.

evidential images. Masks, glasses, beards, aging (**Fig. 7**), and variations in weight cause problems for both manual and automatic methods.

Image analysis is further used for face reconstruction techniques.

Many commercially available systems use biometrics data to verify a claimed identity. For instance,

Figure 6 A less successful enlargement of an image from a surveillance video by bilinear interpolation due to low spatial resolution of surveillance system.

coded iris data stored at smart cards can be used to authenticate a cardholder's identity. These methods are not fully applicable for the same reasons as for automatic face recognition.

Object recognition

Object recognition defines the process of detecting and identifying objects to determine if a physical item is identical to an object in an image. For this purpose visual methods, superimposing digitized images or metrology-based methods are used. As for 3D facial identification, creating a 3D volume of the physical exhibit can simplify the comparison, although it is not commonly used at this time.

Investigations often concern vehicles from videotapes. The shape, texture and color are used to determine class similarities and individual features such as dents and scrape marks, are used for identification. Methods for automatic interpretation of license plates are commercially available, based on image processing and pattern recognition methods. This requires high quality images. Otherwise, image-

Table 3 Methods for person identification

Visual comparison	
Comparison using microscopes	
Metric comparison, manual or semiautomatic measurements of the facial features and geometrical distances	Absolute measurements Relative measurements
Computer-based methods for 2D facial images	Perspective alignment 2D superimposition techniques
Automatic face recognition algorithms	Statistical methods Learning based systems
Computer-based methods for 2D and 3D facial image comparison	
Computer-based methods for entirely 3D facial images	

processing techniques can be applied to enhance the visibility of the characters, though not always successfully, mainly because at low spatial resolution.

Image measurements

Image measurements can provide information about the course of events and the size of items and persons in a crime scene. The height of a person might for example add information useful in an investigation. Measurement techniques are based on photogrammetry and 3D-reconstruction and are successfully used to extract data from 2D images. 3D-reconstruction techniques are also used for crime scene documentation and analysis. The choice of methods depends on the parameters available; fixed or mobile camera position; fixed optics or zoom; known or unknown place of photography/video recording; if there is a single image, sets or sequences and if there are one or several camera views.

As an image is only a 2D representation of a real scene, additional information is needed for measuring in the scene. One way is to use calibrated cameras and metric values to provide the data needed, but calibration requires a visit to the crime scene. Known item sizes in combination with geometric image correction techniques can be used in the case of noncalibrated cameras.

Reverse projection is a useful technique for single camera systems that require fixed cameras and optics, which surveillance systems normally provide. An additional recording is required where the same recording system and camera position is used, for scaling the measured values into metric values. This also takes into consideration any lens distortion. By applying this method to an image sequence, the uncertainty of the measured height due to various body postures can be reduced (**Fig. 8**).

The 3D-reconstruction technique using calibrated cameras provides rather accurate measurements, if the place of the crime can be located. This technique is based on the use of more than one camera view and a calibrated camera to obtain complementary data. Metric values are also needed for scaling. In practice, the scene is reconstructed by calculating and comparing the distances from objects in the scene, the image

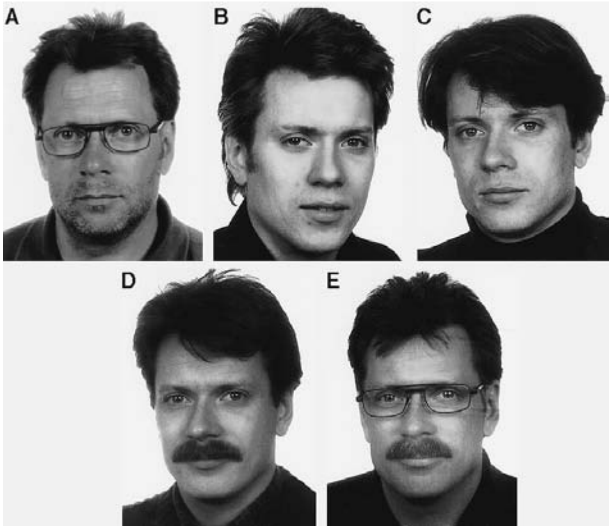


Figure 7 Variations of a person's face over time. **A**, June 1999; **B**, 1975; **C**, 1986; **D**, 1994; **E**, 1995.

Figure 8 Measuring the height of a robber by reverse projection from a digitized surveillance video recording. The height was calculated from an image sequence and the path of the robbers movement in the bank have been tracked.

planes and the camera positions. In the case of variable focal length or unknown camera position this method is preferred.

Some methods allow for the use of only noncalibrated cameras. In this case there must be well-known item sizes in the scene for scaling of the measurements. Additional correction techniques are often used in combination.

The difference between measuring simple rigid objects and the complex measuring of the nonrigid human body is well worth noting as accuracy strongly depends on the person's position and dynamic behavior. Shoes, clothes and masks further affect measured values.

Technical investigation of image storage media and recording devices

In many cases there is a need for tracing the image source to determine a suspect's connection to a crime scene. Methods for identifying recording devices, to determine copying, recording and editing order of a video recording and to determine the time and place of production of the videotape might clarify such connections to evidential images. The recording devices leave traces in the image storage medium why storage medium itself provides a set of useful features. These methods can for example be used to find out whether a specific video camera has been used to record child pornography videotapes or in order to trace technical alteration.

For this purpose the image information can be analyzed as well, to determine if and how the image content or image sequences have been altered. Deviations and variations in video signals, irregularities on charged coupled device (CPP) sensor surfaces, variations and noise patterns caused by the recording mechanism in a video camera can give some clues. For digital devices there is sometimes additional information in image file headers, data files, or time codes supplied by cameras. The audio signals can be used to trace the type of video recording device, for example audio format and number of audio channels.

From the physical features of the videocassette information such as manufacturing, time of production and place of production can be determined by using the printed information on the cassette houseing, box and labels. The same holds for CDs, although there might be more information in the data blocks assigned to manufacturing information. The magnetic tracks can be used to determine the authenticity of a video recording or to determine copying, recording and editing order in a similar way as for magnetic audiotapes.

Electrical signals provide further information about the recording devices, and methods for trouble shooting television and video are useful. A commonly used method is to analyze the head switch position (**Fig. 9**), which can be used to identify recording devices and to determine the copying order. Other signals are the dihedral error, that is the position of the video-head on the drum of the video tape recorder, the position of the CTL-pulse versus the video signal, the RF-envelope and the FM-carrier of the luminance signal. These signals can be measured for videotapes and compared to the corresponding measures from a questioned video tape recorder.

Characters inserted in the video sequence can also be useful information. The characters differ in functionality, size, shape and position among cameras, which holds for conventional and digital still cameras as well as for conventional and digital video cameras.

Related techniques

Images in large databases, e.g. a digital reference library, as well as in combined databases are today

Figure 9 A horizontally and vertically delayed picture where the head switch position can be seen.

essential for rapid and automatic search. Such databases can contain reference libraries of child pornography. The automatic search engines are based on features defined by the image texture, color and object size for example. The algorithms are often statistical or learning-based and might also provide rapid on-line search.

Image transfer has become an important area, as time in many cases is a critical parameter. Telecommunications for information exchange, such as image transmission to a remote database for on-line comparison of the images in the database, is often needed.

Scientific Areas

The pixels in a digital image are not randomly ordered as are the picture elements in a photograph (**Fig. 10**). Each pixel is addressed to a certain color and position in the image, and to time in an image sequence. The number of pixels in the image related to the true size of the object is referred to as image resolution. Pixels can further be used to define clusters of pixels, where contours, shapes, colors, texture and pixel motion are used to define how the pixels are related. These pixel clusters further provide for analysis of the various items in the depicted scene by semantic analysis. Adding the time, not only analysis of single images but also image sequences, which extend the dimensionality. These are the basic low level features from which the scientific areas have evolved into higher level analysis.

Image processing is normally referred to as enhancing poor images by creating images that are easier to interpret. Lately semantic analysis has expanded image processing into new exciting areas.

Image analysis refers to the science of extracting information from images, although values other than images are the output. The aim can also be to reduce the large amount of data for computational reasons. The output can be numerical, for example the object positions in an image. The analysis can be on a low level where each pixel is treated separately, or on higher levels, semantic or context-based image analysis.

Figure 10 Picture elements in a photo versus pixels in a digital image.

Computer vision can mainly be seen as the science of modeling imaging systems for processing and analysis. Computer vision differs from image processing and analysis in the way that the goal is to analyze the scene content, rather than just the 2D-image information.

The basic goal of *Image coding* is to deal with efficient ways to represent images for transmission and storage purposes involving data compression. Image compression techniques are based on the fact that our eye is rather insensitive to minor changes of geometry, contrast and resolution in an image. As this information can be considered redundant, it can also be reduced without seriously affecting the visual quality of an image. Examples of image compression techniques are the various JPEG and MPEG standards and methods based on wavelet or fractal coding schemes.

Computer graphics are often referred to as the process of generating images. These methods can successfully be used for image alteration in combination with image processing. Computer graphics are used for simulation, illustration, reconstruction and documentation.

Photogrammetry is the process of measuring from images that provides techniques for measuring a 3D scene from 2D images. This field can be combined with computer vision, image processing and analysis for reconstruction of a scene and the objects in the scene.

There are two major subgroups of automatic image *interpretation methods*: statistical and learning-based techniques. Statistics can be defined as the identification of structure in an image, and the handling of uncertainty. Statistical interpretation often requires *a priori* knowledge, such as face geometry in the case of automatic identification of an individual from facial images. Learning-based systems require sets of images for training purpose. Methods, such as artificial neural network and fuzzy logic, fall into this category. Common to all automatic recognition methods is that the variations between the classes or individuals to be identified must be larger than variations within the class. The result strongly depends on the choice of representation model.

Pattern recognition can be used to detect, recognize and identify objects, automatically based on the mentioned interpretation methods. Methods used in biometrics, for example, aim to fully automate interpretation of image features such as faces, iris patterns and fingerprints.

A common subject for discussion is 'Who tells the truth, the eye or the technology?' This question is especially relevant for forensic image analysis. The human visual system is enormously complex, and

superior to any computer vision system, for example in dealing with motion and dynamic changes. In spite of this, there are areas where the computer can be used as a complement to the eye: to enhance the visibility (**Fig. 11**), and provide numerical values. There are also ways to fool the brain. This makes perception of visual information an important field for dealing with evidential images, even if the definition of *neuroscience* is still in the making. The study of the brain will most likely pinpoint and provide solutions to the problems with visual interpretation of images, focused on the areas of visual systems, coupled to behavioral and cognitive neuroscience.

The process of exploring, transforming and viewing data as images, referred to as *visualization*, can be used to illustrate the result of image processing or to plot numerical values from an analysis. Image storage, image transmission, database search engines, data fusion, sensor fusion and signal processing are other related areas of importance, as well as physics, optics, electronics and computer science.

Conclusion

Dealing with visual evidence requires methods for extracting all possible information, as the image quality is frequently very low. Evidence from other methods, such as DNA or fingerprints, often yield a better result. In some cases the evidence consists of only a worn out, noisy video recording, and the outcome of forensic image analytical methods might then add clues to the identification process. These methods can also be used for cases that could not have been solved by traditional methods or would require considerable time and resources.

Videotapes from surveillance system recordings are the dominate image source and, no doubt, the major problem source. Currently, the best way of achieving image enhancement and restoration is to improve surveillance system technology. Even if the enhancement techniques are continuously developed, the information lost at the imaging stage is impossible to fully recover.

No doubt images will remain important evidence, although the storage medium will probably alter by bringing images closer to the more general concept of 'digital evidence'. It is obvious that there is further need for highly specialized techniques for image analysis that can help in the daily work in forensic laboratories.

See also: **Anthropology**: Overview. **Computer Crime**. **Forgery and Fraud**: Payment Cards, **Crime-scene Investigation and Examination**: Recovery of Human Remains; Scene Analysis and Reconstruction. **Document Analysis**: Analytical Methods. **Evidence**: Statistical Interpretation of Evidence/Bayesian Analysis. **Pattern Evidence**: Bare Footprint Marks. **Photography and Digital Imaging**: Digital Imaging Enhancement. **Facial Identification**: Photo Image Identification; Computerized Facial Reconstruction; Skull-photo Superimposition. **Electronic Communication and Information**.

Figure 11 Variations of resolution and contrast.

Further Reading

Bear MF, Connors BW and Paradiso MA (1996) *Neuroscience, Exploring the Brain*. Baltimore: Williams & Wilkins.

Brady M (1981) *Computer Vision*. Amsterdam: North Holland.

Bruce V and Young A (1998) *In the Eye of the Beholder, The Science of Face Perception*. New York: Oxford University Press.

Fukunaga K (1989) *Statistical Pattern Recognition*. New York: Academic Press.

Granlund GH and Knutsson H (1995) *Signal Processing for Computer Vision*. Dordrecht: Kluwer Academic.

Howard IP and Rogers BJ (1995) *Binocular Vision and Stereopsis*. New York: Oxford University Press.

Jähne B (1997) *Digital Image Processing, Concepts, Algorithms, and Scientific Applications*, 4th edn. Heidelberg: Springer.

Katani K (1996) *Statistical Optimization for Geometric Computation: Theory and Practice*. Amsterdam: Elsevier.

Klasén L (1999) *Image Analysis*. In: Frank RS and Peel HW (eds) *Proceedings of the 12th Interpol Forensic Science Symposium, Coordinating Laboratory Reports on Evidence Types*: October 20–23, 1998. pp. 261–302.

Kraus K (1993) *Photogrammetry*. Bonn: Dümmel Verlag.

Schäfer R and Sikora T (1995) Digital video coding standards and their role in video communications. *Proc IEEE* 8:907–924.

Sonka M, Hlavac V and Boyle R (1993) *Image Processing, Analysis and Machine Vision*. Cambridge: Chapman & Hall.

Torres L and Kunt M (1996) *Video Coding The Second Generation Approach*. Boston: Kluwer Academic.

Watkinson J (1994) *The Art of Digital Video*. Oxford: Focal Press.

Digital Imaging Enhancement

G Oxlee, Kalagate Imagery Bureau, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0599

Introduction

In the overview article on photography and digital imaging, Lena Klasen stressed the importance that images and imagery have played in forensic science. In recent years the importance of imagery has been underlined by the rapid increase in the number of closed-circuit television security and surveillance systems that have been deployed, not only in high-risk installations, such as banks, retail outlets and key government establishments, but also in city and town centers. In whatever field of endeavor, the key to the successful use of imagery is the ability of trained imagery analysts to extract useful data from the imagery. The initial extraction of data is, of course, through the eyes, but these are often aided by enhancement devices such as spectacles and optical lenses, and so on. However, it is the human brain that provides the object recognition and the analysis of what part the object plays in the scene. The analyzed data are then compiled with other data to provide useful 'imagery intelligence'.

The brain's capability to interpret images can be enhanced by training to provide better understanding of the way in which energy is reflected from an object, recorded by the imaging system and subsequently released as a picture for the eye to observe. This process is more complicated than it at first seems, and without imagery analysis training observers often come to the wrong conclusions. However, even experienced imagery analysts sometimes need electronic help. There are two main factors involved here. First, our eyes (as remarkable as they are) are far less sensitive than the optical and electronic devices that record and replay the imagery; second, a large part of the imagery available to forensic imagery analysts is of relatively poor quality. The fact is that there is information in the imagery that the eyes cannot 'see' and it is thus necessary to enhance the picture to the point where these data come within the accommodation of the human eye. Because this encyclopedia covers forensic work, this article first briefly covers some of the legal constraints surrounding digital imagery in general and digital enhancement in particular.

Imagery Enhancement: Legal Constraints

Digital and optical enhancement

Although historically it was initially seen as a dubious new science, photography has been accepted in the courts for many years. Photography is understood by most people in the court and a photograph or negative is easily seen to be first evidence. Video too has gained rapid acceptance and is wrongly seen as moving photographs, albeit of poorer resolution. However, digital imagery, and in particular digital enhancement, is viewed with suspicion by many courts of law and has not yet been fully accepted. The effects, frequently demonstrated in the cinema and on television, of both optical and digital enhancements are impressive, convincing and exciting but

they are effects that are known to be often overwhelming falsehoods. Thus, it is essential that automatic safeguards are in place before enhancements are accepted for forensic purposes.

Nothing must be extracted or, more importantly, introduced into the original imagery record, be it optical or digital. The original is the first evidence and must remain intact and unsullied. Any subsequent enhancements can be compared with this original to test their authenticity as evidence and their probative value. Moreover, all specific enhancements produced for the courts must be processed on equipment that has an audit trail, so that the results can be replicated by other forensic imagery analysts.

Many of the enhancements that can be produced by digital means could be done using analog optical equipment but optical methods are both time consuming and costly. Digital enhancement therefore predominates, and will continue to do so.

Data compression

There are a vast amount of data in a single picture. Each picture element within a digital image is known as a pixel, and even in a simple scene there can be many millions of pixels. The exact number will depend upon the resolution of the recording apparatus and the quality of the electronic peripherals that support it. Because of this vast amount of data, there is a need to compress the data for many stages of the process, starting from scene being imaged and recorded and on through to the final picture to be analyzed.

Data compression is described in detail elsewhere. Among other things, its use has the enhancement advantage of reducing unwanted noise. There are many data compression systems and algorithms, and hybrid versions appear almost daily, but the fundamental principle that must be observed in forensic work is that the system used must not be predictive. To provide industry standards, a group of photographic experts set up a committee to lay down a basic standard and to judge the merits of various algorithms. This decompression standard was called JPEG (Joint Photographic Experts Group Image Compression Format). The algorithms for JPEG are effectively ‘lossless’ because all of the data on each field is sampled. Thus the system is nonpredictive.

The JPEG system, with its large data sampling rate, proved to be cumbersome and unwieldy for moving video pictures. Thus a system called MPEG (Moving Picture Expert Group) was devised. MPEG samples the data in each incoming picture less frequently. It recreates an intermediate frame (picture) by a prediction process. Basic MPEG is thus not suitable for forensic use. However, subsequent developments (called variously MPEG2 and JMPEG) have built-in safeguards and many are now suitable for forensic work. The key is that they must not be predictive in their nature. In simple (and possibly oversimple) terms, the system must not guess (predict) and then insert this predicted data into the reconstructed frame.

Even though a compression system is nonpredictive it does not make it automatically suitable for forensic work. The effectiveness, in terms of the ability to compress and subsequently reproduce a picture that is as near to the original as possible, is inter alia dependent upon additional factors. The refresh rate of ‘I’ frame data from an original uncompressed frame must in any event be high, in order to prevent data drift and more and more errors occurring.

There are other elements involved in compression but these can mostly be considered as subsets of the base algorithms designed to increase the efficiency. Wavelets is but one example; there are many others. Wavelets was devised in the 1940s before the complexities of digital Fourier transforms were fully developed. Thus Wavelets has been a successful part of decompression for many years but it is important to restress that these and any other subsets are only suitable for forensic imagery work if the overall system is fundamentally nonpredictive.

Authenticity of digital images

Superimposed upon any decompression or recording system is the need to guarantee the authenticity of the imagery. Strict rules are laid down in most justice systems on the ownership and handling of imagery data. The history of any digital imagery evidence must be available to the court and authenticated by each person in the procedural chain.

Digital imagery, unlike analog imagery, does not have a clear and obvious original. With a conventional film camera the negative is the original and the untouched negative would thus be the first evidence in court. However, with imagery that is recorded digitally, the stream of data is passed to a processor and it is a computer chip that recovers the data initially and stores to a hard disk. In essence, therefore, it is the hard disk that is the original. Nevertheless, a great advantage of digital storage and enhancement is that *all* copies are replicas and are thus indistinguishable from the original. Analog systems, on the other hand, produce copies with consequent data loss, albeit normally quite low and not noticeable to the naked eye.

In regard to digital systems, the alternative to handing the court a negative is to have special authentication for the original. Two examples of how this can be done are watermarking and encryption.

Watermarking An electronic watermark can be imposed within the frame data. Thus picture ownership can be authenticated. The main problem is that, since the watermark will always persist even after enhancement and other manipulation, it would be a target for 'hackers'.

Encryption Modern digital systems have sophisticated encoding that make interception of the signal difficult and effectively prevent anyone from reproducing an unscrambled and thus meaningful picture. Of course, without this base picture anyone with 'Machiavellian' intentions could not manipulate the imagery. Tampering with the picture is thus impossible without access to the encryption keys. It is, however, important to ensure that encryption is integrated early enough in the process to preclude signal alteration before the initial recording is completed.

Imagery Enhancement Techniques

As established above, enhancement is necessary to reveal to the imagery analyst data that would, in their raw recorded state, be invisible or imperceptible to the human eye. The results of any enhancement are dependent upon the quality of the original imagery. The nature and quality of the image will largely dictate the type of enhancement tools used. A typical enhancement menu on a modern workstation will have facilities to enlarge, sharpen, contrast-stretch, blur, smooth and detect edges. It is not be possible to cover all of the filters in the space available: a representative few are discussed below.

Enlargement

There are several methods of digital enlargement. Four examples are pixel expansion, near or natural neighbor, bilinear interpolation and cubic convolution.

Pixel enlargement

This is the simplest method of enlargement. The area of the picture accommodated by each single pixel is made larger. The problem with pixel enlargement is that the picture becomes a series of squares, which disguises the true nature of the scene after relatively few pixel expansions. When dealing with images with a great density of pixels giving very high resolution, the pixels can be expanded further, but even then the breakthrough of squares limits the degree of enlargement. An example of pixel expansion is shown in **Figs 1** and **2**.

Figure. 1 is an aerial photograph of a collection of farm buildings. The area is typical of a rural scene of crime. At this scale it is difficult for the eye to see individual buildings. Optical enlargers could be used but, as the picture has been scanned, digital enlargement is also an option. Figure 2 is an example of pixel expansion. The buildings are larger but the interpretability of the picture is marred by the presence of the pixel squares. This is called pixel breakthrough. The greater the expansion, the more distorted become the features in the picture. This fact is used in some television broadcasts when the face of an individual must not be recognized for legal reasons and is deliberately subjected to pixel breakthrough. It is quite an effective method of concealment but not very safe because an enterprising technician could manipulate the pixels and recover the facial details. Thus, more sophisticated methods of blanking facial data are generally used.

Figure 1 Aerial photograph of a collection of farm buildings.

Near neighbor or natural neighbor

Some but not all of the pixel distortion can be lessened by using near neighbor interpolation. This is sometimes called natural neighbor. The interpolation uses the pixel closest to the pixel location of interest. This smoothes the most dramatic impact of pixel expansion, and in a sense averages out the

Figure 2 Pixel expansion of farm buildings shown in Fig. 1.

distortion. However, even using near neighbor algorithms, pixel breakthrough distortion is still a problem when the expansion factor is large.

Bilinear interpolation

Bilinear interpolation algorithms are based on a weighted average of the four pixels surrounding the pixel location of interest. With this system of enlargement the pixel breakthrough is all but completely eliminated.

Cubic convolution

Cubic convolution is an older and more complex interpolation algorithm than bilinear interpolation. However, it too is based on a special averaging of the pixels surrounding the pixel location of interest. With cubic convolution, pixel breakthrough is eliminated and the system maintains the integrity of the object with respect to its background, irrespective of the degree of enlargement. The limitations are therefore the same as those of optical enlargement lens systems, i.e. the degree of enlargement is limited only by the resolution of the picture.

Figure 3 is an enlargement of the same farm buildings seen in Figs 1 and 2. This time the enlargement has been achieved using cubic convolution and there is a clear absence of pixel breakthrough.

Figure 4 is a scanned oblique photograph typical of the sort of picture taken from a police helicopter. It shows the same farm. **Figure 5** is an enlargement of the farm tractor and trailer that was located in the centre of Fig. 4. Cubic convolution has also been used for this enlargement and once again there is a clear absence of pixel breakthrough.

Figure 6 is yet further enlargement of the tractor tire area and it still shows no pixel breakthrough, although the resolution limits of the picture become clear.

Figure 4 Scanned oblique photograph of farm buildings shown in Fig. 1.

Sharpening

There are several filters that have the role of sharpening the image. These include edge detection and edge sharpening.

Edge detection

There are numerous filters that enable edge detection. This is essential for edge sharpening and other enhancement processes. Among the main filters are the series of Laplacian edge detectors. These filters produce sharp edge definition and can be used to enhance edges with both positive and negative brightness slopes. All types of Laplacian filters work on a weighting system for the value of the pixels surrounding the pixel of interest. In a 3×3 filter the system

Figure 3 Enlargement of farm buildings (Fig. 1) using cubic convolution.

Figure 5 Enlargement, using cubic convolution, of tractor and trailer shown in Fig. 4.

Figure 6 Further enlargement of tractor tires shown in Fig. 5.

diagnoses the eight pixels around the main pixel. The system weights to maintain a balance where the sum of all the weights equals zero. On or near the edges of the image, the values of edge pixels are replicated to provide enough data. Another edge detector is the Sobel filter. This too is an omnidirectional spatial edge enhancement filter and also uses a 3×3 matrix to calculate the 'Sobel gradient'.

Figure 7 is a well-defined image of a fingerprint. Resolution and coverage of this sort would create no problems for the fingerprint expert. The introduction of a Sobel edge filter produces even clearer edges and gives the impression of a three-dimensional surface (**Fig. 8**).

Edge sharpening

These filters use a subtractive smoothing methodology. An average spatial filter is applied which retains the frequency data but reduces high frequency edges and lines. The averaged image is subtracted from the original image to leave the edges and linear features intact. Once the edges are identified in this way, the difference image is added to the original. This method provides clearer edges and linear features but has the disadvantage that any system noise is also enhanced. An example of edge detection and sharpening can be seen in **Fig. 9** which shows two pictures of the same farm buildings. The one on the left is the enlargement by cubic convolution. The one on the right has undergone edge detection and edge sharpening. This does not make the picture esthetically more attractive but improves its interpretability.

Figure 7 Fingerprint image.

Figure 8 Fingerprint (shown in Fig. 7) reproduced using edge filter.

Contrast Stretching

It is frequently necessary to stretch the contrast of a picture. The most important reason for this is that a picture often contains more data than the eye can accommodate. A full gray-scale monochromatic photograph will have up to 256 gray scales; however, the human eye can only 'see' up to 26 of these. Contrast stretching will enhance subtle differences between the object and its background by bringing the contrast difference into the accommodation range of the human eye. Contrast stretching can also aid the eye when analyzing color imagery. Human perception is vastly improved when color is introduced. It is said that the eyes can 'see' up to many millions of tones of color. It is thus often advantageous to portray a monochromatic scene in color. This is usually

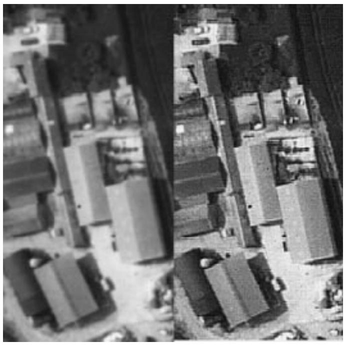


Figure 9 Farm buildings shown in Fig. 1: enlargement by cubic convolution (left); result of edge detection and edge sharpening (right).

false color but if the 256 gray scales in a black and white photograph are changed to color then there is very little difficulty in seeing the subtle gray scale differences.

In **Fig. 10** the shapes of, for example, the materials in the trailer are difficult to see. However, a simple false color treatment, as seen in **Fig. 11**, reveals the shapes more clearly. In general when more colors are introduced the shapes will become clearer (within the resolution limits of the original imagery).

Contrast variations can be controlled in many ways but for most PC-based systems the mouse is the most accommodating device. Variations can be made either throughout the picture or in specific parts. In **Fig. 12** part of a fingerprint cannot be seen clearly, while the other part is usable. By applying contrast stretch filtration and mouse manipulation to the area selected, the hitherto unclear fingerprint is revealed (**Fig. 13**). Enlargement further assists the eye.

Contrast filters

There are a broad range of filters that will provide contrast stretching. Many of these use histogram operations. Histogram operations can tone down the brightness response curve. This has the effect of altering the distribution of contrast within the spectrum of dark to bright pixels.

Figure 10 (see color plate 46) The materials in the trailer are difficult to see.

Figure 11 (see color plate 46) Color treatment reveals the shapes in Fig. 10.

Figure 12 Part of this fingerprint is usable but part cannot be clearly seen.

Histogram operations include histogram stretch. This operation expands the response curve in a linear manner. Contrast is thus spread evenly throughout the picture or area of interest. The bright ranges are not competing against one another but images with large areas of dark and large areas of bright pixels are not strongly affected.

Another histogram operation provides histogram equalization. This operation modifies the response curve nonlinearly. The total pixel dynamic range is used in a balanced and uniform manner. This emphasizes the contrast in the bright areas, as in these areas the contrast is increased the most.

Histogram operations also allow the operator to change from a positive image to a negative image. In complex and difficult images the eye can sometimes appreciate contrast better in the negative domain, and vice versa.

Color contrast can be modified by filters that impact upon the tonal transfer curve. There is normally a set of curves for each family of colors. The curves each allocate a shade of gray in images with particular intensities. Once the tonal transfer function is applied, the images are displayed with a specific color contrast.

Figure 13 Contrast stretch filtration, manipulation and enlargement enhances the image shown in Fig. 12.

Blur

Sometimes it is advantageous to reduce the amount of detail in order to better appreciate the overall picture. In general, forensic imagery analysts take a complex scene and simplify it so that particular information only is seen. In this they are rather like thematic mapmakers who put on to a map only that which is helpful to a particular subject or theme. This is to make the data more digestible. For example, if a person is provided with a photograph or series of photographs of New York State and asked to navigate in a car from New York City to Albany on minor roads, he or she would find the task very daunting. This is because there is too much detail in the photographs to make them suitable for the job. However, if provided with a road map on which there is only the basic data required, the task becomes easy. Thus it is that in some cases data on a photograph have to be minimized. Blur filters can be used to produce the desired effect.

A blur filter is a form of edge smoothing. Its use reduces the prominence of high contrast spots and edges. It can therefore reduce clutter, which would otherwise cause distraction, and will also help to reduce system noise.

A common blur provider is the Gaussian filter. It uses the Guassian function G(i,j) = exp (−((i−u)**2 + (j−v)** 2)/(2* SIGMSQ)) where (i,j) is a pixel in the filter 'window', (u,v) is the geometric centre of the window and SIGMSQ is set to 4. The system is set so that the sum of all the weights is 1.

Smoothing

Some smoothing is achieved with blur filters but it can also be achieved using a median filter, when pixels of interest are replaced using a weighted median value derived from its neighbors. It computes the median values usually within a 3×3 window of pixels surrounding the pixel of interest. When the median or middle value is applied, it has the effect of smoothing the image while preserving the key edge data.

Noise Reduction

By definition, noise is unwanted data. Noise reduction filters can be applied at any stage of the enhancement. This is particularly important as many enhancement techniques enhance the noise as well as the required signal. Noise reduction can be achieved in a number of ways.

Morphological functions contract or expand the edges and borders of uniform light or dark regions of a picture. The contraction function replaces each pixel of interest with its darkest neighbor. This causes light objects to contract minimally and considerably reduces light-colored signal noise. The expansion function replaces each pixel of interest with its brightest neighbor, causing dark objects to shrink minimally and dark-colored signal noise to be dramatically reduced.

An average filter, sometimes called a mean filter, is also used to smooth pictorial data to reduce noise. It carries out spatial filtration on each pixel, again typically using a 3×3 window. The filter calculates the total sum of the value pixels within the window and this total is divided by the number of pixels in the window. The values of edge pixels that are near to the borders of the image are replicated to provide enough data for filtration in the edge regions.

In specialized imagery, such as active microwave systems, gamma filters are used to remove high frequency noise, which is called radar 'speckle' (while preserving high frequency edges), or, in the case of infrared, special versions are used to reduce system saturation. Such filters can be used to reduce small-scale system saturation on passive infrared imagery. In essence, the gamma filter carries out spatial filtration on each individual pixel, using the gray level values in a square matrix surrounding each pixel. The matrix can be from a 3×3 format to an 11×11 format.

Warp, Rotation and Roaming

Using a technique called warp, an oblique image can be portrayed in a different perspective from the original perspective that was recorded by the camera. It does this while maintaining the original scale and aspect ratios. This is a complex operation which rotates the geographical coordinates of all points of the picture. It can be useful in determining the relationship of an object to its background or in superimposing one image upon another when the angular field of view of the second image is slightly different from that of the first.

Rotation algorithms allow the picture to be turned on its central axis. Thus, for example, an oblique photograph, taken at an angle to the horizontal and where the vertical objects (e.g. the side of a building) are portrayed at a slant, can be rectified.

Roaming allows the operator to roam around different parts of a large image and to follow a line of interest from one point to another.

Virtual reality modeling

The power of modern computers has enabled many complex operations. These in turn have provided an

interesting form of digital imagery enhancement. On very poor forensic imagery, if the dimensions of specific objects in the scene have been derived from accurate on-site measurement, then a three-dimensional computer-generated model can be produced. This computer model can be superimposed upon the photograph to show more clearly what the photograph itself portrays. Moving objects, such as persons within the scene, can themselves become modeled and the whole scene can be rotated into any perspective required. The operation is complex and costly and there are those that are concerned about its safety as an evidential tool. However, providing safeguards such as watermarking the model, encryption of the data and audit trail of the process are put in place, fears should be allayed.

Stereoscopy

Stereoscopic imagery is an important form of enhancement. Apart from holographic images, imagery is a two-dimensional representation of a three-dimensional world. Imagery analysts, along with most humans, are used to viewing life in three dimensions. It follows, therefore, that if imagery can show the third dimension (depth), its contents are more easily identified and analyzed.

A three-dimensional image is obtained because we have two eyes, each looking at the same scene from a slightly different angle. The brain is able to fuse the two images and create a three-dimensional scene.

Most imagery taken for forensic work appears as a single image, either in photographic or in video form. But if a scene of crime operator were to take two pictures of the same scene or object, one after the other from slightly different viewpoints (in azimuth not range), then three-dimensional viewing is possible. The processed images are viewed through special glasses called stereoscopes. These allow the left eye to look only at the left picture and the right eye to look only at the right picture. The brain then carries out its normal fusion function and the three-dimensional scene is apparent.

Some work has been done using video to produce a stereoscopic effect. In this case the camera is stationary but the subject moves. Providing the movement is in the horizontal (azimuth) plane, then a three-dimensional effect is possible. Neighboring stills are used and observed through stereoscopes.

Unfortunately most closed-circuit television systems in static situations, e.g. in banks or city centers, are time-lapse systems. That is to say that the output of several cameras is recorded on to one tape and the tape is slowed to last for 12 or 24 hours. Therefore, often too few of the 50 fields per second produced by the camera are recorded and the movement of an object (for example a robber) between recorded frames is too great for the three-dimensional effect to be possible.

Stereoscopic imagery can be taken from a digital camera system or, if analog, can be scanned and digitized. Once in digital form, each frame can be placed electronically at the correct (customized) interocular distance for viewing.

An on-screen three-dimensional system uses two digital stereoscopic frames. These can be polarized: one horizontally and the other vertically. Each eye lens of viewing spectacles is polarized in correlation and hence, when the screen image (images) is viewed, the three-dimensional scene is apparent on the monitor.

Conclusion

Digital imagery enhancement is a vital tool in forensic imagery analysis. However, safeguards are important if the dramatic impact of imagery on crime investigation and the prosecution of criminals is not to be abused. Watermarking, encryption and audit trail are key to the elimination of public concern about digital imagery and digital imagery enhancement.

There are many filters to enable a wide variety of electronic enhancement, and more are being developed. Nevertheless, the key that unlocks the data in an image is the ability of the human brain to provide analysis based on observation and common sense. This ability can be enhanced by learning and experience.

Pictures can be misinterpreted and thus mislead. Trained and experienced imagery analysts are required if serious errors are to be avoided in the field of forensic imagery analysis. The eyes of even trained imagery analysts are vastly inferior in performance and sensitivity to the electronic sensors that image and record a scene. Digital enhancement of imagery to bring the imagery data within the range of human eye performance is therefore often a prerequisite to successful imagery analysis – where the brain functions in its full analytical mode.

See also: **Photography and Digital Imaging**: Overview.

Further Reading

House of Lords Select Commitee (1998) *Digital Images as Evidence*, Ch. 2 Background; Ch 3, Digital Images, London: HMSO.

Jahne B (1996) *Digital Image Processing*. Heidelberg: Springer.

Schafe R and Sikora T (1995) *Digital Video Coding*. *IEEE Proceedings* 83(6):907–924

Sonka M, Hlavac V and Boyle R (1993) *Image Processing Analysis and Machine Vision*. Cambridge: Chapman and Hall.

Watkinson J (1994) *Art of Digital Video*. Oxford: Focal Press.

Polygraph *see* **Detection of Deception**. **Lie Detection**: Polygraph.

POSTMORTEM EXAMINATION

Procedures and Standards

P Saukko, University of Turku, Turku, Finland
S Pollak, University of Freiburg, Freiburg, Germany

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0778

Short Historical Overview

Anatomical dissection

The prerequisite for the whole idea of an autopsy or necropsy was the knowledge of human anatomy, and the history of autopsy shows three, partly overlapping, paths of development. The first of them lasted almost 2000 years, from the first known school of anatomy in Alexandria around 320 BC, where human dissections were carried out, until the publication of the great textbook of anatomy *De Humani Corporis Fabrica* in 1543. Written by Andreas Vesalius (1514–1564), the 'father of anatomy', this marked the overthrow of traditional galenic anatomy. The possibility of dissecting human bodies varied greatly at times. In Vienna the first anatomical dissection took place in 1404, and in Prague somewhat later in 1460. Felix Platter I, the famous anatomist in Basle, was said to have performed more than 300 autopsies since 1559.

Medicolegal autopsy

The medicolegal investigation of deaths was introduced relatively early as a result of the requirements of the judicial system. According to Singer, the earliest of medicolegal dissections took place at the University of Bologna, Italy between 1266 and 1275. In France, Ambroise Paré performed the first medicolegal autopsy in 1562.

Our knowledge of old autopsy procedures is rather scanty. With few exceptions, detailed written autopsy records are relatively recent. Such exceptions are the reports of the autopsy of Emperor Maximilian II (1576) and of Markgrave Jakob III of Baden (1590).

The principles of modern medicolegal investigation were based on the codes of sixteenth century Europe: The Bamberg code (*Constitutio Bambergensis*) in 1507, the Caroline Code (*Constitutio Criminalis Carolina*) in 1532 and later the Theresian Code (*Constitutio Criminalis Theresiana*) in 1769.

The Austrian decree of 1855 contains very detailed instructions, in 134 paragraphs, as to the performance of medicolegal autopsy, and is still legally valid today in Austria. Similar, although not as detailed, is the Prussian edict of 1875. Both of these instructions can be considered as the culminating point of legislation dealing with the performance of medicolegal autopsy.

Clinical autopsy

The clinical autopsy, as we understand it today, took much longer to develop and first became meaningful after the introduction of modern concepts of pathogenesis of disease, by Carl von Rokitansky (1804–1878), and cellular pathology, by Rudolf Virchow (1821–1902).

At the beginning of the nineteenth century increased attention was paid to the actual autopsy technique. Prost, a French physician, insisted in 1802 that all organs of the body should be examined, and declared that 3 hours was the minimum time for a postmortem examination. In 1846 Rudolf Virchow, then prosector in Berlin, insisted on regularity and method and definitive technique. The classical techniques, which are still in use today, are more or less modifications of those introduced by Rokitansky, Virchow, Ghon and Letulle, among others.

In 1872 Francis Delafield's *A Handbook of Postmortem Examination and Morbid Anatomy* was published in New York, and German and English editions of Rudolf Virchow's book on autopsy technique were published in 1876.

Present Use of the Autopsy

Medicolegal autopsy

Further development of medicolegal autopsy has been characterized and greatly influenced by the judicial system adopted in any given country, the

main emphasis being in the detection and investigation of criminal and other unnatural or unexpected deaths. Due to different legislation and practices, there exists great variation in medicolegal autopsy rates between the countries. In addition to national measures to create guidelines and to harmonize medicolegal autopsy, there has been increasing international interest in achieving harmonized and internationally recognized rules on the way autopsies should be carried out. This has become imperative, especially from the point of view of human rights. The mass killings in Cambodia, Rwanda, Bosnia and Kosovo should have made it quite clear, even to the general public, what implications a medicolegal investigation, or the lack of it, may have upon human rights.

In May 1989 the UN Economic and Social Council adopted, in its resolution 1989/65, the Principles on the Effective Prevention and Investigation of Extra-Legal, Arbitrary and Summary Executions, that had been created by cooperation with intergovernmental and nongovernmental organizations, especially the Minnesota Lawyers International Human Rights Committee. In 1991 the General Assembly of the United Nations endorsed the Model Autopsy Protocol of the United Nations.

The European Council of Legal Medicine (ECML) is an official body that sits in Cologne and deals with scientific, educational and professional matters at a European level. It has delegates nominated by the national medicolegal associations from all European Union and European Economic Space (EES) member countries. Since the early 1990s the ECLM has also been active in this field and its document *Harmonisation of the Performance of the Medico-Legal Autopsy* was adopted by the General Assembly in London in 1995.

The Council of Europe is an intergovernmental organization whose aims among others are to protect human rights and pluralist democracy. It should not be confused with the European Union. The two organizations are quite distinct; however, all the 15 European Union States are also members of the Council of Europe, which currently has a total of 41 member states. In its 43rd Ordinary Session, the Parliamentary Assembly of the Council of Europe adopted a Recommendation 1159 (1991) for the harmonization of autopsy rules. Following this recommendation, a working party of international experts in legal medicine and law, with representation from Interpol and the International Academy of Legal Medicine, was established in 1996 under the Committee of Bioethics to make a proposal for the Autopsy Rules. One of the guidelines used in the work was the Autopsy Rule produced earlier by the ECLM. The working party finished its work in November 1998 and this new paneuropean recommendation No. R (99) 3 on the harmonization of medicolegal autopsy rules and its explanatory memorandum was adopted by the Committee of Ministers on 2 February 1999 at the 658th meeting of the Ministers' Deputies. Although the document is a 'recommendation' by nature, and hence strictly speaking not legally binding, it has legal implications because all 41 Council of Europe member countries have agreed to incorporate these principles in their national legislation.

Clinical autopsy

Despite the advent of more sophisticated investigative and imaging techniques, the clinical autopsy has been shown to have maintained its value and remained an essential factor in the quality assurance of medical care. Regardless of this, there has been a progressive decline in autopsy rates throughout the world. The mandatory 20% autopsy rate required for accreditation of postgraduate training in the United States was withdrawn in 1971, on the grounds that each institution should set its own rate but that it should ideally be close to 100%.

According to World Health Organization (WHO) statistics published in 1998, total autopsy rates varied in Europe in 1996 between 6% (Malta) and 49% (Hungary) and, in other parts of the world reported to the WHO, between 4% (Japan) and 21% (Australia) (**Table 1**).

The reasons for this decline are many and complex: overreliance on the new diagnostic techniques, low appreciation of autopsy work, poorly performed autopsies by inexperienced trainees without proper supervision, long delays in the production of autopsy reports, economic factors, fear of malpractice litigation, etc., to name just a few.

The standardization and harmonization of clinical autopsy has taken place primarily at national level. The quality of health care and quality assurance and audit have become increasingly important and have been extended from laboratory medicine to encompass autopsies.

Objectives of Autopsy

An autopsy is a detailed systematic external and internal examination of a corpse, carried out by a pathologist or one or more medicolegal experts, to ascertain the underlying and possible contributing causes of death and, depending on the jurisdiction, also the manner of death. Before the pathologist can begin the examination he or she must be sure that the

Table 1 Reported information on autopsy rates 1996

Country	*Autopsy rate (%) all ages*
WHO region of the Americas	
Canada	20
USA	12
WHO European region	
Albania	Not available
Austria	27
Belgium	Not available
Bulgaria	25
Czech Republic	31
Denmark	32
Finland	36
France	Not available
Germany	8
Greece	Not available
Hungary	49
Iceland	38
Ireland	7
Israel	Not available
Italy	Not available
Luxembourg	Not available
Malta	6
Netherlands	8
Norway	9
Poland	9
Portugal	Not available
Romania	7
Spain	Not available
Sweden	37
Switzerland	19
United Kingdom: England + Wales	24
United Kingdom: Northern Ireland	11
United Kingdom: Scotland	15
WHO Western Pacific region	
Australia	21
Japan	4
New Zealand	16
Republic of Korea	Not available
Singapore	16

autopsy has been authorized. An assessment of possible risks that may be involved with the autopsy must be considered, and necessary health and safety precautions taken. The autopsy and all related measures must be carried out in a manner consistent with medical ethics and respecting the dignity of the deceased.

An autopsy is performed to achieve one or more of the following objectives:

- To identify the body or record characteristics that may assist in identifying the deceased.
- To determine the cause of death or, in the newborn, whether live birth occurred.
- To determine the mode of dying and time of death, where necessary and possible.
- To demonstrate all external and internal abnormalities, malformations and diseases.
- To detect, describe and record any external and internal injuries.
- To obtain samples for any ancillary investigations.
- To obtain photographs or retain samples for evidential or teaching use.
- To provide a full written report and expert interpretation of the findings.
- To restore the body to the best possible cosmetic condition before the release.

In addition to the anatomical dissection, there are basically two main types of autopsy:

- The clinical autopsy is carried out to investigate the extent of a known disease and the effectiveness of treatment, and it is sometimes also performed for medical audit or research purposes. Almost invariably the consent of relatives is needed unless the deceased has given consent ante mortem.
- The medicolegal or forensic autopsy, which is ordered by the competent legal authority (a coroner, medical examiner, procurator fiscal, magistrate, judge or the police) to investigate sudden, unexpected, suspicious, unnatural or criminal deaths. Unidentified bodies or deaths occurring in special circumstances, such as deaths in police custody or during imprisonment, are also often subjected to a medicolegal autopsy. In most jurisdictions permission of the relatives is not required.

Autopsy Techniques

Both clinical and medicolegal autopsy may involve different strategies and techniques, depending on the questions they are expected to answer. Autopsy technique in adults is generally somewhat different from pediatric autopsies.

The scope of medicolegal autopsy is much often broader than that of clinical autopsy and may also include the investigation of the scene of death. All background information on the circumstances of death are of paramount importance in choosing the right approach. In medicolegal autopsy the examination of the clothing is often an essential part of the external examination, whereas in clinical autopsy it is generally not. Both types of autopsy should consist of full external and internal examination of the body, including the dissection and investigation of all three body cavities.

External examination

External description of the body includes the age, sex, build, height, ethnic group and weight, nutritional

state, skin color and other characteristics of the deceased, such as scars or tattoos; description of postmortem changes, including all essential details relating to rigor mortis, hypostasis and decomposition; careful investigation and description of all body surfaces and orifices including color, length, density and distribution of hair, color of irises and sclerae, presence or absence of petechiae or any other abnormalities or injuries. The examination should be carried out systematically and include head, neck, trunk, upper and lower extremities and the back.

Internal examination

Examination of the body cavities includes the description of the presence of gas (pneumothorax), fluids (effusions or exudates) or foreign bodies and the measurement of their volume, appearance of the internal surfaces and anatomical boundaries as well as location and external appearance of organs.

The classical autopsy techniques vary mainly in the order in which the organs are removed:

- The organs may be removed one by one (Virchow technique).
- Cervical, thoracic, abdominal and pelvic organs can each be removed as separate blocks (Ghon technique).
- They may be removed as one single block, which is then subsequently dissected into organ blocks (Letulle technique).
- All organs are dissected *in situ* (Rokitansky technique)

All organs have to be dissected, the outer appearance as well as the cut surfaces described and the weight of major organs recorded. The hollow organs have to be opened and their content described and measured. All relevant vessels, arteries and veins as well as ducts have to be dissected. All abnormalities must be described by location and size.

Sampling

Histological examination of the main organs should be performed in all autopsies. The need for further ancillary investigation may depend on whether the cause of death has been established with the necessary degree of certainty, and, if not, additional samples have to be taken for toxicological or other investigations. For toxicology this may include peripheral blood, vitreous humor, cerebrospinal fluid, bile, hair samples or other relevant tissues. When retaining tissues one has to take into consideration possible restrictions, depending on national legislation.

Special procedures

Sometimes special procedures and modifications of normal dissection techniques are necessary. Chest X-ray has to be performed before the autopsy if there is suspicion of air embolism. Where neck trauma is suspected, the brain and the organs of the chest cavity have to be removed before the dissection of the neck to drain the blood from the area to avoid artifactual bleeding. Postoperative autopsies may present various problems with medicolegal implications, such as complications of anesthesia, surgical intervention or postoperative care. Detailed description of these special dissection procedures and techniques is beyond the scope of this presentation.

Autopsy Report

The report is an essential part of the autopsy. It should be full, detailed and comprehensive. Medicolegal autopsy reports, in particular, should also be comprehensible to the nonmedical reader. In addition to the factual, positive and negative gross, microscopical and analytical findings, the pathologist should conclude with a discussion of the significance of the findings. Where the findings are of uncertain nature and there are several competing causes, the pathologist should try to give an opinion as to their probability.

See also: **Legal Aspects of Forensic Science**.

Further Reading

Council of Europe (1999) *Recommendation No. R (99) 3 of the Committee of Ministers to Member States on the Harmonisation of Medico-Legal Autopsy Rules*. Strasbourg: Council of Europe.

King LS and Meehan MC (1973) A history of the autopsy. *American Journal of Pathology* 73:514–544.

Ludwig J (1979) *Current Methods of Autopsy Practice*, 2nd edn. Philadelphia: WB Saunders.

Morris M (1990) *Report on the Harmonisation of Autopsy Rules*. Strasbourg: Council of Europe.

World Health Organization (1998) *World Health Statistics Annual 1996*. Geneva: Geneva:WHO.

Post-mortem Interval *see* **Entomology. Time Since Death**.

Preservation *see* **Crime-scene Investigation and Examination**: Preservation.

Profiling *see* **Geographical Profiling**.

Psychiatry *see* **Psychology and Psychiatry**: Overview.

PSYCHOLINGUISTICS

D Canter, Centre for Investigative Psychology,
University of Liverpool, Liverpool, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0781

Written and Spoken Utterances as Special Evidence

Spoken or written threats, suicide notes, confessions, declarations in wills, and a range of other utterances may become part of civil or criminal proceedings or the associated investigations. These reports of what people have said or copies of what people have written may be examined in order to answer a number of different questions: for example, exactly what the utterances mean; what is likely to happen as a consequence of them; what the characteristics of their author are likely to be; who the author is likely to be; or whether they are genuinely the product of the person who it is claimed produced them.

The value of studying words that are written as part of a crime is that text does appear to offer special potential as evidence. A threatening extortion letter is a record of exactly what the offender did to commit the crime. The letter is the crime scene. There is no need to make sure the witness saw what was happening or to puzzle over the pathologist's report. On the page is a complete record of the relevant criminal actions.

Similarly, a note left declaring the reasons for a suicide is a valuable piece of evidence, if found with a dead body. It may help to show what the person was thinking and feeling and thus lend support to any belief that there were no other culprits. It is a very special kind of record of inner despair that has no equal in observations of what the person did before death, or even what others report he or she had said.

The written confession is also held in high regard by the courts and public alike. It is the fact that culprits have described in their own words the actions that incriminate themselves that is so significant. Their words are seen as providing direct contact with their guilt.

There are many other aspects and types of crime in which some form of written record of the words of the offender exist. Stalking may be one such, in which offensive letters may play an important part. Various types of business fraud may also leave traces of the writing of the offender. All these written accounts offer a direct glimpse into the mind of key actors, if we can find a way to interpret those glimpses.

In some cases the questions are about the usual or particular meaning and usage of the words. An example would be whether the utterance was likely to be interpreted as a genuine threat by the person who received it, or whether the text implies connection with a brand in a trademark dispute. In these cases the expertise required to answer the questions draws upon knowledge of the language in question and would be provided by a linguist, typically one who has taken a particular interest in the forensic application of linguistics.

Psycholinguistics Distinguished

When the questions about written or spoken material deal with aspects of the characteristics, behavior, intentions and mental state of the author, the

expertise involved is more likely to draw upon that branch of psychology that deals with linguistic issues, often known as psycholinguistics. Most notably, for instance, is the consideration of whether a suicide note is genuine, as might be part of a psychological autopsy. However, the overlap between the different disciplines and pseudodisciplines that have an interest in this area is quite considerable. Often differences in expert opinion on the same verbal material will be a consequence of differences in the discipline that is brought to bear, rather than some substantive difference in actual measures or findings. Because of these potential confusions some more clarifications are of value before we turn to consider the modest contribution that psycholinguistics can currently make to forensic matters.

The first clarification to make is that the pseudodiscipline of 'graphology' has no place in the forensic process. Graphologists claim to be able to interpret the character and mental state of an author from the way he or she shapes letters on paper when writing. Graphology therefore focuses on the actual physical way in which the author writes and how they arrange the words on the page, in order to make psychological inferences about the writer. Although graphologists are widely used in industry, especially in France, for selection of personnel, and there are many anecdotes about their contributions to police investigations, every thorough systematic, scientific study has failed to produce evidence to support the claims of graphology. Furthermore, graphologists claim they must know the writer's age, sex and nationality before they can comment on the personality because none of these facts is revealed by the writing. For many investigations, of course, these crucial facts are what the investigator wants to know.

There are also some basic problems in using the analysis of handwriting: although handwriting does consists of measurable elements, such as slant and size, and descriptive elements, such as letter form and tendencies to the right and left, there has been little progress in establishing a common denominator for any of these elements to permit comparable objective evaluation. So although there is still a scientific interest in the possibilities that graphology may offer, there is no consistent evidence of its validity. Nor is there any psychological reason to expect that personality should be revealed through the ways in which written letters are formed. And, of course, in many situations criminals deliberately avoid writing. The widespread use of word processors is doubtless further reducing the impact of this pseudoscience.

Graphology should not be confused with the work of handwriting analysts. These forensic scientists often have a training in one of the natural sciences, typically physics. Their task is to examine the physical structure of any written material to determine whether it has been modified, or for comparison with other examples, for instance to see if a signature has been forged. They do this in much the same way as fingerprint experts, by determining the salient features of the material in question and establishing what proportion of the features are common to any target and comparison material. It is also important to separate what might be called the physics of writing analysis. Whether it is the shape of a signature or the font of a typewriter, it may be possible to carry out precise analysis to link the physical trace to its originator. But no-one carrying out such analysis would suggest that it could also tell you whether the writer was paranoid or likely to follow through with his or her threats. When the language exists in the form of sound rather than writing, questions about its authenticity or similarity to other material will draw on knowledge of phonetics, requiring the expertise of a forensic phonetician.

Stylistics

Some attempts to determine the authorship of written material have drawn very heavily on numerical procedures. The number of words with more than five letters in the text have been counted, for example, or the frequency with which certain combinations of words occur, such as 'and then'. There is a tendency to refer to this highly numerical approach in determining whether a piece of text is in the style of one author or the other as 'stylistics'. The term is sometimes rather loosely applied to any approach to establishing the distinct qualities that determine the particular characteristics of an author.

These numerical methods have become more common with the advent of computers and may draw heavily on the skills of linguists in determining parts of speech and unraveling the structures of language. Historically, however, this area has suffered from being used in contexts in which the actual author of the text is unknown. The procedures are then used to suggest authorship, even though there is no external validation. It is on the basis of such explorations that claims have emerged that it is possible to determine the authorship of a text, or at least whether or not it has multiple authors. However, when carefully controlled experiments have been conducted using texts of known single and multiple authorship, they have been unable to support claims that any stylistic procedure can reliably distinguish single from multiple

authors or indicate who an actual author is likely to be.

The most notorious of these 'stylistic' procedures is known as 'Cusum' or 'Qsum', standing for cumulative summation. This procedure, developed by Reverend Quentin Morton, uses arithmetic procedures derived from engineering quality control. In essence it consists of finding the average number of two-and three-letter words per sentence in a piece of text and graphing the cumulative differences from this average across the full text. On to this graph is imposed a further graph, derived from the cumulative differences from the average of the total number of words in each sentence. Morton and his associates claim that any apparent divergence of these two graphs from each other is a prima facie indication that different authors are involved in the creation of the text. Such a curious claim was allowed into a few court cases until systematic research showed that it had no empirical validity.

Contributions of Psycholinguistics

In some cases relatively straightforward psychological considerations of texts can be of great use. For example, it has been found in a number of cases of anonymous, offensive letters that the author reveals so much about his or her own knowledge and intellectual capabilities through characteristics in the way the letters were written that it was clear that the person continually described in the third person could only have been the writer. This type of revelation is more likely when a fluent but poorly educated writer is too arrogant to see how he or she is giving him- or herself away. In most cases the examination of what is written is much more difficult, so much so that a whole new science is struggling to take a first few steps, gleaning whatever can be of investigative value from the examination of written texts.

The questions that forensic psycholinguists are seeking to answer fall into three groups. One set is to do with the identity of the writer. Is this the way in which a known individual expresses him- or herself? A second set of questions relates to the character of the writer, really trying to answer the same types of questions as graphologists about the traits and personality of the person writing. The third falls under the heading of prognosis. What is the writer likely to do under various circumstances?

Identifying Authorship

The questions of identity have probably been the most explored and have certainly been the ones that have most exercised police officers. Claims that confessions are not the words of the accused have opened the doors to a series of appeals against conviction and complaints against the police. In these cases the forensic linguist or psycholinguist is claiming that the accused has a unique style of expression and that the confession contains indications that are inconsistent with that style. This draws on the rather attractive notion that we all have a unique style of speaking or writing. The task of the analyst is to specify the details of that style and then demonstrate whether the questioned text accords with those details.

A few studies do show that, under certain, special circumstances, specific indicators may act as 'signatures', such as very specialist vocabulary, jargon or slang, unusual spellings or inaccuracies in grammar, or peculiar ways of putting the text on the page, punctuating it or adding emphasis. These may be of use as evidence when two documents are being compared; however, it is difficult to find any examples in the legal record where such linguistic analysis has been accepted without question, although it may still assist investigations.

One example that illustrates this approach is a case in which an incriminating, but anonymous, diary was compared to the known writing of a suspect. A number of misspellings were found in both the questioned and known texts (e.g. 'breath' instead of 'breathe', 'its' instead of 'it's'). These also had consistencies in the form of error that could be related to how the words might be pronounced. There were also a number of profanities common to both sets of text, such as 'ass', 'butthole' and 'screwed'. Furthermore, there were similarities in the way time was recorded and the forms of expression of emotion. A number of grammatical constructions were also shown to be similar in both samples. It was therefore proposed that the two sets of writing had the same author.

But, even in such an apparently strong example, there are problems in being certain that the indicators really can be treated as 'signatures'. It is possible, for instance, that most people in the suspect's circle of associates misspell 'breathe' and 'it's', that 'ass', etc. are common words in their vocabulary and that they all express themselves in a similar way. Without knowing the prevalence, and co-occurrence, of all these constituents in the relevant community, such matches can only be taken as a useful indicator.

Any consistent differences that might be found between the writings of two people may well be a consequence of the language community from which they come. It might be, for example, a product of a particular school, or prison, not unique to an individual.

Even musicologists cannot distinguish all Haydn from all Mozart because both these composers use many of the idioms of their time, as we all do in language.

The problems of knowing the prevalence in the population of any particular aspect of language may make it difficult to use those habits as evidence of identity, but it does not rule out the possibility of narrowing the search for suspects by knowing which language subgroup a writer might come from. Those who study the social differences in the use of language, sociolinguists, have explored the different ways in which people communicate in different subgroups of the population; for example, the larger vocabulary and more complex sentence constructions of educated people and the tendency of professionals to use passive verb forms are widely recognized.

As attractive as the notion is of us each having a unique style of written or spoken expression, or there even being distinct subgroups in the population who have identifiable styles, many processes reduce the possibilities of each person having an identifiable style in all their communications. Verbal communication is a means of contact with others. It is therefore modified in a host of subtle ways to adjust both to the content of what is being expressed, and to whom it is being expressed. Police officers will certainly express themselves in a virtually different language when telling a bawdy joke to friends in a pub compared with when giving evidence in court. We all tend to use much richer and denser vocabulary, and more grammatically correct forms of language, when writing than when we speak. As a consequence, the measurement of a person's style has somehow to be calibrated to the context in which the communicating is taking place.

Studies have shown, for example, that people use rather different grammar and vocabulary when writing extortion letters compared with when they write personal threatening letters. Some aspects of their mode of writing may carry across the two contexts, but any simple measure of style, such as vocabulary or syntax, turns out to be surprisingly dependent on the nature of the threat being expressed. Much more careful study is therefore needed before we can be confident of clear stylistic differences between people in how they write. This requires the study of the frequencies with which words, groupings of words and grammatical forms occur in different contexts. Developments in computing are now making this possible by the analysis of vast collections of examples of text. But these bodies of material at present tend to be very general, drawn often from newspapers and other published sources. To make a real breakthrough in understanding texts of interest to the police, we need to work on samples of the relevant types of documents.

At a rather more general level of analysis it may be possible to detect similarities in *what* a person writes rather than *how* it is written. This is really an exploration of the person's ways of thinking rather than the style of expression. Such studies are open to many difficulties in obtaining precise measurement and objective indices, but it has been shown, for instance, that genuine suicide notes have a distinctly different psychological tone to those that are written as simulations for experimental purposes. Genuine notes tend to be longer and more explanatory, indicating more clearly that the author has internalized the decision to take his or her own life.

In a similar form of analysis it has been found possible to determine if two extortion letters are by the same person. They may be typed in very different ways, with intriguing differences in vocabulary, but an examination of the themes in the letters and how they are organized can sometime indicate remarkable similarities. For example, both may start with an instruction to take the threat seriously, followed with a declaration that the threat was feasible, give specific instructions about how the money should be bundled, repeated warnings about not involving the police, and so on. These similarities of approach may indicate common mental processes suggesting a common author.

Inferring Characteristics of Authors

Given the difficulty in pinning down unique aspects of personal style, it is not surprising that attempts to link writing style to personality and other characteristics has been even less successful. The difficulty does seem to be the need, somehow, to take account of the particular contexts in which the utterances occur. For example, even threats may be expressed differently, depending on the type of threat being expressed. People are likely to reveal their traits in quite different ways when trying to bully an associate into keeping away from their girlfriend compared with when they are trying to extort money out of a supermarket chain. Following this line of thought, there has been moderate success in recent studies. For example, people who used more profanities in personal threat letters did have personality profiles that showed a greater desire to control other people. But clearly there is a long way to go before this line of study can be precise enough to be really useful to a police inquiry.

The possibility for distorting the modes of expression in written material, trying to hide the characteristic style of the writer, is another factor that always needs to be considered. But these attempts may reveal more than they hide. It is difficult for unsophisticated writers to emulate sophisticated ones, but people who have fluency and proficiency find it difficult to totally hide that. In one murder case a suspect had kept a careful diary in which her visit to the victim was carefully recorded in the most benign, casual terms. However, the suspect had been so obsessional in the way she kept the diary that it was clear she would not normally have recorded such a passing visit in the way she did. A careful analysis and count of all the references in the diary and how they were handled demonstrated that the key entry was unusual in some striking ways, in part because the text it contained was so unremarkable. This analysis was drawn on by the prosecution counsel to shape his crossexamination on the diary, but never presented as evidence. The suspect was convicted of the murder.

Predicting Consequences

Language can be analysed at many levels of complexity: the number of particular types of word, the patterns of certain combinations of words, the forms of grammatical structure used, the types of idiom and the themes that are expressed. Even the approach taken to the task of communication may be examined, how carefully planned it is, what prior knowledge it implies, and so on.

It is the higher level, more general forms of analysis of themes and approach that may be especially useful for answering the questions about what the writer is likely to do; how is he or she likely to react to different forms of intervention?

The majority of people who write threatening letters never follow through on their threat. The act of writing is the expression of anger or frustration, malice or spite. Against that backdrop, the task, then, is to detect those letters that reveal a determination to act. Some explorations do suggest that a number of aspects of the themes and approach to the writing of threats can be used as criteria to predict action. It is not appropriate to publish these in a public forum, but it is worth mentioning that they draw upon a careful analysis of the credibility of the threat and the benefits and costs to the writer of carrying out the threat.

In terms of strategies for dealing with threats, police forces therefore need to distinguish carefully the exact nature of the promised threat. Quite different issues and modes of analysis will be appropriate for individualized threats for personal revenge, political threats on ideological grounds and extortion for money, although one type of threat may masquerade as another. Meticulous study of the form of words in which a threat is expressed can be of great value in understanding what is likely to be characteristic of writers' forms of expression, what they are really trying to achieve, what sort of people they are likely to be, and the probability of their doing anything other than writing such letters.

Conclusion

In conclusion it has to be emphasized that, despite the claims of a few academics, there is still no strong evidence that a person's writing style can be uniquely pinpointed, measured and used for comparison with contested examples. There has been no legal case in which forensic psycholinguistics has been accepted without question to establish or disprove the identity of a defendant. It seems most unlikely that we will ever establish unique 'fingerprints' in written or spoken styles of communication. The way we express ourselves in words is just too flexible to allow that degree of precision. However, although its use in court may be limited, there are many ways in which careful psychological and linguistic analysis of the crime scene of a written document can help investigations.

In all other aspects of forensic psycholinguistics the analyses at present tend to be rather *ad hoc*, depending on the very particular circumstances of the utterances. Whether there will ever be the possibility of a generic process for answering any of the forensic psycholinguistic questions is still a matter of debate.

See also: **Psychological Autopsies**. **Phonetics**. **Investigative Psychology**.

Further Reading

Aked J, Canter D, Sanford A and Smith N (1999) In: Canter D and Alison L (eds) Approaches to the attribution of authorship. *Profiling in Political Practice*, pp. 157–188. Aldershot: Ashgate.

Canter D (1992) An evaluation of the Cusum stylistic analysis of confessions. *Expert Evidence* **1**:93–99.

McMenamin GR (1993) *Forensic Stylistics*. Amsterdam: Elsevier.

PSYCHOLOGICAL AUTOPSIES

D Canter, University of Liverpool, Liverpool, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0782

Introduction

From time to time incidents of suspicious death occur in which the mental state of deceased persons needs to be assessed. If some evaluation can be made of the sort of person they were, their personality and thought processes, especially as that may throw light on any involvement they themselves had in their death, then it may assist the investigation of what is sometimes referred to as 'equivocal death'. Such an evaluation, known as a psychological autopsy, is an attempt to reconstruct a person's psychological state prior to death.

The typical case for which a psychological autopsy may be of value is one in which there is some doubt as to whether death was accidental, self-inflicted or malicious, and whether the deceased played an active role in his or her own demise. Such matters can be especially important in life insurance claims that are void if the death were suicide. They are, of course, also of potentially great significance in murder inquiries, such as those in which there is a question as to whether the deceased contributed to his or her own death in some way. Fatal accident investigations, in which the technicalities of what actually led to the accident are difficult to resolve, are other forms of equivocal death in which psychological examination of aspects of the main actors/victims may be essential.

Virtually all attempts to use psychological procedures to throw light on a person's thoughts and feelings prior to their death have taken place in the USA, most of them in civil and criminal litigation rather than as part of an investigation. It has been claimed, for example, that 'the drawing of a psychological picture of a dead person whom the expert has never met' can be a 'tool for criminal defense attorneys'. The expert who draws such a picture can be used in court to argue, for example, that a person accused of murder was acting in self-defence, or whether a gift is in contemplation of death and therefore of relevance in considerations of taxes due on death.

Many equivocal death examinations, however, have been part of civil proceedings in which a worker's compensation case makes it necessary to establish that certain events affected the deceased in a particular manner, or that work-related injuries contributed to the eventual suicide of an individual. Another example would be where a will is contested, so the mental state of the deceased is the major legal battle ground. In the case of Howard Hughes, for instance, a psychologist concluded that 'psychological problems, numerous head injuries, and drug misuse had changed a vibrant millionaire into an emaciated recluse'. The possibly psychotic basis of his reclusiveness, rather than mere eccentricity, posed challenges to the probity of his estate.

A possibly less obvious circumstance in which the characteristics of a person may be usefully inferred, even though that person is not available to answer personal questions, is an inquiry into the report of a missing person. Another example is when it is necessary to understand the decisions made, and actions taken, by people who are no longer able to answer for themselves. For example, in the planning of care routines for demented patients it is inevitable that some inference will be made about the mental state of the people in question. Making those inferences as systematically as possible, drawing on whatever psychological information is available, is likely to improve the decisions being taken.

Psychological Autopsies

Attempts to assess the mental state and characteristics of a person who is not available for direct examination contrast with the task that usually faces mental health professionals: effective mental health assessments typically require the participation and cooperation of the examinee. Although in some circumstances the examinee may choose not to cooperate, it is considered essential that the examinee be available for evaluation. Furthermore, the current mental state and adjustment of the individual concerned is the focus of attention, notably in competency and capacity evaluations such as child custody, competency to plead and competency to testify examinations. For even if subjects of assessments are not willing to participate they can be directly observed, their demeanour considered and their transactions with others noted and other aspects of their daily lives recorded directly. If they are not present, all these aspects of them, which can assist in assessing

their mental state, have to be inferred at second or third hand.

Even though evaluations conducted in the absence of the examinee are fraught with challenges and problems, this has not stopped a number of clinical psychologists, psychiatrists and law enforcement agents in the USA from producing such assessments. They have given them a variety of terms, including 'psychological autopsies', 'psychiatric autopsies', 'reconstructive psychological evaluations' or 'equivocal death analyses', even 'offender profiles' in some cases.

Useful distinctions can be drawn between these different activities, depending on whether the target of the examination is actually known. In the case of the examination of a crime scene to infer characteristics of the offender, often referred to as 'profiling', the main quest is to determine the identity of the perpetrator. But where the identity of the subject is known, as in the attempt to determine whether the death was by accident or suicide, the objective is to reconstruct the mental state of the deceased. If this can be based upon information obtained from people who had direct or indirect contact with the subject of the examination, then it is most likely to be called a 'psychological autopsy'. In cases where the investigation is carried out by law enforcement officers, usually FBI agents, who only examine the crime scene material and other information directly available to the police inquiry, it may be called an 'equivocal death analysis' or EDA. Clearly an EDA is open to many more biases and distortions than a 'full' psychological autopsy. What little experience there is about EDAs casts grave doubts on their validity.

The difference between a literary autopsy, or a military character analysis, and a psychological one are likely to be differences of degree rather than kind. A psychologist would be expected to provide more systematic detail and to give clearer evidence for the conclusions reached. It might also be expected that the psychologist would more confidently express views on the motivations and personality of the target than would other authorities. However, the contribution of psychologists to the investigation of equivocal deaths is still very limited, and probably not as different from the work of a literary biographer as some psychologists would like to think.

Perhaps the most obvious difference to hope for, between the literary recreation of a bygone celebrity and a psychological autopsy, is that the psychologist would draw upon what is known of people and processes similar to that demonstrated in the actions at the centre of the inquiry. The individual being considered would be taken as an illustration of people and processes who are known to have carried out similar actions. In this way, the expert opinion would be clarified and bolstered by the empirical evidence of other known cases. For example, knowledge of how adolescent suicides prepare to take their own life, or what is typically contained in a suicide note, may be drawn upon to develop the account of the individual and his or her actions, before death.

Unfortunately, there is still very little detailed empirical evidence available on many topics that are relevant to contributing to equivocal death investigations. Perusal of the literature also indicates a lack of a comprehensive assessment and evaluation of the nature and validity of those psychological autopsies that have been carried out. It is therefore most appropriate to consider the psychological autopsy as a relatively unstructured technique, in which mental health professionals attempt to describe the thought processes and personality of a deceased person prior to death, and in some cases to comment on their likely participation in their own death. Its contribution to equivocal death investigations may therefore be best regarded as the development of an organized framework for indicating the issues to be considered when forming a view about the deceased.

The most common equivocal death scenarios that are examined in the USA are those in which suicide is suspected but is not absolutely certain. The first people to contribute to such examinations are generally regarded as being Shneidman and Farberow at the Los Angeles Suicide Prevention Center during the 1950s. They responded to requests from the coroner that they assist in determining the cause of death in equivocal suicides. Information from persons related to the deceased was combined with the coroner's (i.e. pathological) findings to determine the cause of death as either suicide or accident. Shneidman and Farberow gave technical definition of the psychological autopsies that they carried out for the coroner: 'A retrospective reconstruction of an individual's life that focuses on lethality, that is, those features of his life that illuminate his intentions in relation to his own death, clues as to the type of death it was, the degree (if any) of his participation in his own death, and why the death occurred at that time'.

The Technique

As has been indicated, contributions to equivocal death investigations can range from an essentially informal attempt to reconstruct the thoughts of the deceased to a much more thorough exploration of everything that is known about him or her. For, although such contributions have been made over a period of approximately 40 years, there are still no standardized procedures that have been agreed upon for making them. However, where more systematic

methods are employed, they have commonly involved obtaining information from interviews with survivors of the deceased and archival sources. Shneidman and his colleagues interviewed relatives, friends, employers, physicians and others, including teachers and in some cases even bartenders, who could provide relevant information in an attempt to reconstruct the deceased's background, personal relationships, personality traits and lifestyle. They sought significant details of the events immediately preceding the death. All of this information was subsequently reviewed by the death investigation team in the coroner's office, resulting in a determination of the mode of death.

Shneidman subsequently developed an outline for conducting a psychological autopsy, which essentially consisted of a 16-point check list (Appendix I) that is not that dissimilar to the framework that might be used by any physician in preparing a medical case history. The major differences from other forms of medical case history are the focus on what is known about the deceased's typical actions, especially reactions to stress, what might be known of the person's interpersonal relationships, thought processes and experiences surrounding the death. The procedure is based on the assumption that people close to the deceased can provide accounts of both historical and recent developments and behaviors of the deceased. Likewise, historical and recent archival information, such as physicians' records, may contribute to a determination of the individual's mental state at a particular point in time. The procedure also implies a form of corroboration in which as wide a range and variety of sources of information are collected to ensure that the bias inherent in any one source of information does not distort the whole picture.

The problem of bias is an especially important one, given the importance that may be given to the psychologist's opinion about the cause of death. Yet few writers on this process discuss it in any depth. Litman and coworkers are possibly the most direct in drawing attention to recurring problems. They point out that there is often a lack of information about the individual, particularly information that could be used for a reliable inference regarding his or her psychological state. Secondly, the information may be distorted by the informants. They cite instances of evasion, denial, concealment and even direct suppression of evidence. Indeed it may be expected that expert advice is required precisely in those situations in which there are doubts and ambiguities surrounding the events of the death. That, after all, is what makes it 'equivocal', so the problems described are part of the reason why an expert is called in to help.

In an attempt to reduce these inherent difficulties Brent proposed that particular attention be paid to the choice of informants, the manner of approach to informants, the effect of the time period between death and the interview on the quality of the information obtained, and the integration of the various data sources. He suggested broadening the range of informants, particularly to include peers in the case of adolescent suicides. He also provided guidelines for approaching informants, emphasizing professional distance and the avoidance of platitudinous commiseration. Brent also reports that he found no simple or consistent relationship between timing of the interview and the quality and quantity of the data obtained. He further recognized that, although the integration of various data sources is a common problem in psychiatry, it is a particularly salient issue in the administration and interpretation of the psychological autopsy. Brent reviewed the relationship between direct and indirect interviews as methods of obtaining family history, suggesting that both sensitivity and specificity of data may be enhanced by increasing the number of informants and including more female informants.

Thus, one of the contributions of behavioral scientists to the formulation of opinions on a suspicious death is the greater care and systematization that they bring to the process, drawing on sources that might not normally be considered by legal professionals or physicians. The most comprehensive set of guidelines, intended to enhance the systematization and move towards standardization of the psychological autopsy, has been provided by Ebert (Appendix II).

Process

In keeping with a procedure that has evolved in response to practical and legal demands, there is no well-developed conceptual or theoretical basis for deriving conclusions from the various sources of information to provide guidance on equivocal deaths. It appears that the professionals involved draw upon their experience to relate the facts to symptoms or syndromes that they would draw upon in their daily practice, searching, for example, for evidence of psychosis, depression or organic dysfunction.

In an attempt to systematize the basis of any guidance, various authorities have proposed some principles, especially for determining if a death was suicide. Perhaps the most clearly stated principle is that expressed by Faberow and Schneidman, that most suicide victims communicate their intentions to others in some way. These principles have been converted into a standardized assessment protocol by Jobes and coworkers. They provide a 55-item Death Investigation Checklist that can be used by medical examiners. On the basis of a successful test of the validity of this

checklist they developed 16 criteria that they called the Empirical Criteria for the Determination of Suicide (ECDS). This process enabled them to clarify determination of suicide as being based on the concepts of 'self-infliction' and 'intention'. This allowed them to derive a score from the ECDS for each of these aspects. By comparing the scores obtained in 35 known accidental deaths and 28 known suicides they were able to show that 'self-infliction' and 'intention' scores both had to be greater than 3 for the death to be declared a suicide.

Although this procedure adds a rare level of precision to what are often difficult medicolegal decisions, the authors emphasize a caveat. Even though, 'medicolegal judgements may be strengthened through such tools, leading to more objective and scientific determination of suicide as a manner of death. It is critical to note that the ECDS instrument is *not* meant to be a rigid and definitive standard designed to usurp the professional's judgement and authority'.

Validity of Psychological Contributions to Equivocal Death Investigations

Given the many difficulties associated with contributing to equivocal death investigations, it might have been expected that there had been many attempts to determine how effective they actually are. Yet there has been very little research that examines the reliability or validity of psychological autopsies or related contributions to inquiries into fatalities. As mentioned, some preliminary work to develop operational criteria for determining suicide as cause of death has been conducted. Fifteen years earlier, in 1976, Shneidman rather disingenuously, referred to the increasing acceptance by coroners of psychologists' opinions as a measure of the validity of those opinions. One group found a high correlation between rank-ordered types of depressive symptoms in depressed suicide victims and in clinically referred depressed patients. This work in effect demonstrated the capacity of interviewers, using a psychological autopsy format, to elicit such information accurately, albeit across a narrow spectrum. Others used the family history interview process to confirm diagnoses obtained via a psychological autopsy procedure, although recognizing that this study was subject to limitations of possible interviewer bias, interviewee bias and a limited sample.

It has been recognized that, even before the validity of psychologists' contributions to inquiries can be sensibly explored, it is important to establish whether different psychologists would offer the same opinion if given the same information. This is the issue of 'reliability' rather than 'validity'. If different psychologists form different views when presented with the same material, then there is little hope that, in general terms, the opinions derived from the processes they used would be correct, or valid.

One study examined the agreement between reports completed by 12 psychologists and two psychiatrists, who reviewed materials that addressed the adjustment and psychological functioning of Clayton Hartwig, suspected of causing an explosion aboard the USS *Iowa* in 1989, resulting in his own death and that of 46 other sailors. Although broad criteria were adopted, only moderate agreement was found between the findings of the 14 professionals conducting the assessment.

This study was initiated as a reaction against an EDA carried out by FBI agents into the USS *Iowa* tragedy. The agents had formed the conclusion that Clayton Hartwig had deliberately caused the explosion in an act of suicide. Their opinion was upheld by the initial Navy inquiry. However, a number of psychologists were highly critical of the EDA and the way it was carried out. Their examination of the facts indicated that the explosion was indeed an accident, a conclusion later supported by the technical evidence and accepted by a US Congress investigation. It was suggested that the FBI used scientific terms for a process lacking significant scientific methodology, and compounded that problem by failing to delineate sufficiently between 'opinions' and 'facts'.

Two critical points must be made. Firstly, as emphasized earlier, the EDA technique was characterized by the examiners not conducting any interviews but relying on information provided to them based on interviews conducted by other parties. Secondly, given the nature of the initial inquiry and the subsequent review, the persons conducting the review had available to them a quantity of information significantly greater than would normally be present for most assessments. Even with this additional information, though, all panellists would have preferred more information regarding Hartwig.

Subsequent to this review, FBI agents who provided the testimony agreed with the view that 'perceived utility and anecdotal evidence are mere proxies for validation, not validation itself'. Yet they rather confused the issue by asserting that EDA was an investigative technique and not a clinical investigative method, somehow implying that the sort of detailed evidence that a 'clinician' might use in helping an investigation was irrelevant to FBI agents trying to achieve the same objectives.

The FBI agents further asserted that EDA is not a clinical but a professional opinion, based on years of

law enforcement experience with indirect assessment of violent death, and the demands of law enforcement require that opinions be provided that do not equivocate. In other words, they are advocating opinions based merely on previous experience without any scientific support for those opinions, or even any possibility for demonstrating that hypotheses alternative to their opinions had been available for test. That they have provided a 'conclusive' opinion in 42 or 45 EDAs, as they claim, is of considerable concern when viewed in the context of the technique's lack of even demonstrated reliability, and the opinion that 'to provide the validity is an exercise in futility'.

In contrast to the surprisingly cavalier view of some FBI agents, the use of blind studies to obtain an estimate of the validity of the psychological assessments of equivocal death victims has been suggested. In relation to the determination of the mode of death, identification of cases in which the 'correct' answer to the question is known has been proposed. Ideally, a group of mental health professionals should be provided with all the information, except for one or two key pieces that clearly identify the correct answer (e.g. a suicide note). The subsequent opinions could then be compared to the 'correct' outcome to provide an estimate of validity. Alternatively, evaluation of 'available' persons and then provision for assessment of all information, except that which is based on the 'availability' of the examinee, has been proposed. Opinions could be compared to those based on comprehensive assessment (including psychometric data) of the individual. Such approaches to address the issues of reliability and validity would seem to be the minimum necessary, but do not yet seem to have been conducted. This suggests mental health professionals should be cautious when using reconstructive psychological evaluations, particularly in legal and quasilegal contexts and those major inquiries, like that in the USS *Iowa* tragedy, in which much of significance to many people may result from the expert's opinion. Professionals are ethically obliged not to mislead their clients about the accuracy of the conclusions they draw.

Admissibility

Despite the many weaknesses of the evidence and procedures used in psychological autopsies and the lack of definitive research to support its validity, it is clear that opinions from psychologists and other mental health professionals, about the mental state of the deceased, have been drawn upon in many courts in the USA and can strongly influence the judgements of those who read them. One experiment showed that 95 medical examiners who received reports containing only physical and circumstantial evidence about a fatality gave distinctly different judgments from 100 who had the same reports augmented with psychological opinion. Given such potential impact, it is perhaps not surprising that many jurisdictions have admitted them.

A significant review of the use of the psychological autopsy in both civil and criminal cases cited the case of *State* v. *Jones*, CR No. 98666 (Arizona Superior Court, 1978), where the psychiatric autopsy of a murder victim was admitted on behalf of the defendant, who claimed self-defense in the killing. The psychiatric autopsy was deemed to show that the defendant, a repeatedly battered wife, was not unreasonable in her assumption that the victim might have killed her. In another case, that of *State* v. *Carrethers*, CR No. 100359 (Arizona Superior Court, 1978), a psychiatric autopsy of the defendant's deceased father was admitted. [In these criminal cases the term 'psychiatric' rather than 'psychological' autopsy was used. Some authorities assert the terms are interchangeable.] The defendant claimed defense of his mother in the killing of his father. The psychiatric autopsy indicated that the father drank heavily, routinely battered his wife and stepdaughter, and provoked numerous acts of violence at work. Again, the psychiatric autopsy revealed that the son had been reasonable in his reaction to his father's violence. Nonetheless, the review highlights the difficulties associated with the admissibility of psychological autopsies as evidence in court hearings, both in terms of Federal and State (USA) rules of evidence and the inconclusive debate regarding the scientific basis for the technique.

There has been a very thorough review of the questions that psychological autopsy evidence needs to answer satisfactorily if it is to be admissible in a US court. Does its probative value outweigh its prejudicial impact? Will it directly assist the trier of fact? Are such autopsies reasonably relied on by experts in the field? Does the proposed witness qualify as an expert? These are exactly the same questions that a British court would wish to answer, but a British court would probably look for slightly different emphasis; for example, perhaps being more concerned about possible prejudice and less concerned about the status of the expert.

It has been pointed out that at least one supreme court in the USA, that of Montana, has directly upheld the admissibility of a psychological autopsy against an appeal that the psychologist should have interviewed the deceased prior to his death in order for his expertise to be valid. The court held that the fact that the deceased 'was not interviewed before his

death does not render inadmissible a psychologist's opinion based on the available data'. However, it was emphasized that, while many courts trying accident claims have accepted psychological evidence concerning the deceased state of mind prior to death, they are more reluctant to permit experts to testify about whether the death was a suicide or an accident.

In applications challenging wills and intestate succession, the courts are reluctant even to accept psychological opinion about the intentions or state of mind of the deceased. This seems to be because such opinion is deemed to intrude too closely into the legal realm. The courts believe that, given the same information available to the expert, they could form a view themselves that would be just as valid. This is not far removed from the opinion that there is indeed much that a layman can do in forming a view about a dead person that will parallel the opinion of a professional psychologist. Furthermore, as was shown in the USS *Iowa* case, there are definite possibilities of prejudice, whereby an opinion presented by an apparent expert is given more credence than it deserves.

All available accounts of psychological opinion relevant to equivocal death cases, used in court, are drawn from the USA. The examples, in the main, though, tend to be civil cases relating to insurance and worker's compensation. Even in the USA, criminal cases are rarely likely to admit psychological autopsies as expert evidence. No instances of its use in courts in other English-speaking countries have been identified. This may, partly, be a function of the data-gathering or terms used. There are clearly circumstances where defense cases, for example in England or Australia, rely on evidence regarding the personality and mental state of a deceased victim in order to promote the issue of self-defense or provocation.

Relevance to Criminal Investigations

The potential for misuse of reconstructive psychological evaluation techniques has been highlighted with reference to the paucity of the research literature and the difficulties illustrated in the Gilfoyle case and the USS *Iowa* inquiries. In using a technique with no well-established reliability or validity, the courts are appropriately cautious about allowing its use in evidence. Without doubt, the value of the procedure can be enhanced by drawing on as many and as varied direct sources of information as possible. A clearly scientific stance, in which alternative hypotheses are tested, would also help to reduce the influence of those agencies or other parties with a biased interest in the outcome of the investigation. Such a stance would require the following:

1. Clear statement of the alternative explanations that are feasible to account for the equivocal death.
2. Clear indication as to the evidence that would be required to support or reject each of these explanations.
3. Full account of the evidence that is available and how it relates to the evidence that would be required.
4. Evaluation of the evidence available and the processes that have been undertaken to test it accuracy and validity.
5. Clear statement of how the evidence has been drawn upon to reach conclusions about each of the explanations offered.

Such a thorough report is very demanding and time-consuming to produce. It also likely to require scholarship and research beyond obtaining the facts of the case and accounts from those close to the deceased. In the pressures of a criminal investigation and the legal process, there may not be all the resources necessary to carry out the task at as high a professional standard as all may wish. In such cases the weaknesses in the psychologist's activities need to be clearly stated, or the professional should actually refuse to undertake the task.

See also: **Clinical Forensic Medicine**: Self-inflicted Injury. **Psychology and Psychiatry**: Overview; Psychiatry; Psychology.

Further Reading

Ault RL, Hazelwood R and Reboussin R (1994) Epistemological status of equivocal death analysis. *American Psychologist* **49**(**1**):72–73.

Barraclough BM, Bunch J, Nelson B and Sainsbury P (1974) A hundred cases of suicide: clinical aspects. *British Journal of Psychiatry* **125**:355–373.

Beck A, Shuyler D and Herman J (1974) Development of suicidal intent scales. In: Resnick K and Leterri D (eds) *The Prediction of Suicide*. Bowie, MD: Charles Press.

Beskow J, Runesow B and Asgard U Ethical aspects of psychological autopsy. *Acta Psychiatrica Scandinavia* **84**:482–487.

Boyd JH (1983) The increasing rate of suicide by firearms. *New England Journal of Medicine* **308**:313–317.

Brent DA (1989) The psychological autopsy: methodological considerations for the study of adolescent suicide. *Suicide and Life Threatening Behaviour* **19**(**1**):43–57.

Brent DA, Perper JA and Maritz G *et al.* (1993) The validity of diagnoses obtained through the psychological autopsy procedure in adolescent suicide victims: use of family history. *Acta Psyciatrica Scandinavia* **87**:118–122.

Canter D (1995) Psychology of offender profiling. In: Bull R and Carson D (eds) *Handbook of Psychology in Legal Contexts*, pp. 343–355. Chichester: Wiley.

Dregne N (1982) Psychological autopsy: a new tool for criminal defense attorneys? *Arizona Law Review* **24**:421–439.

Ebert BW (1987) Guide to conducting a psychological autopsy. *Professional Psychology: Research and Practice* **18(1)**:52–56.

Faberow N and Schneidman E (1961) *The Cry for Help*. New York: McGraw-Hill.

Gould MS and Shaffer D (1986) The impact of suicide in television movies: evidence of imitation. *New England Journal of Medicine* **315**:690–694.

Kovaks M (1985) The Interview Schedule for Children (ISC). *Psychopharmacological Bulletin* **21**:991–994.

Jobes DA, Berman AL and Josselson AR (1986) The impact of psychological autopsies on medical examiners' determination of manner of death. *Journal of Forensic Sciences* **31**(1):177–189.

Jobes DA, Casey JO, Berman AL and Wright DG (1991) Empirical criteria for the determination of suicide manner of death. *Journal of Forensic Sciences* **36**(1):244–256.

Litman RE, Curphey T, Shneidman ES, Farberow NL and Tabachnick N (1970) The psychological autopsy of equivocal deaths. In: Shneidman ES, Farberow NL and Litman RE (eds) *The Psychology of Suicide*. New York: Science House.

Ogloff J and Otto RK Psychological autopsy: clinical and legal perspectives. *St Louis University Law Journal*.

Otto RK, Poythress N, Starr L and Darkes J (1993) An empirical study of the reports of APA's peer review panel in the congressional review of the USS Iowa incident. *Journal of Personality Assessment*.

Paykel ES, Prusoff BA and Myers JK Suicide attempts and recent life events: a controlled comparison. *Archives of General Psychiatry* **32**:327–337.

Poythress N, Otto RK, Darkes J and Starr L (1993) APA's expert panel in the congressional review of the USS *Iowa* incident. *American Psychologist* **January**: 8–15.

Rudestam KE (1977) Physical and psychological responses to suicide in the family. *Journal of Consulting and Clinical Psychology* **45(2)**:162–170.

Selkin J (1994) Psychological autopsy: scientific psychohistory or clinical intuition. *American Psychologist* **49**(1):64–75.

Selkin J and Loya F (1979) Issues in the psychological autopsy of a controversial public figure. *Professional Psychology* **February**: 87–93.

Shafii M, Stelz-Lenarsky J, Derrick A, Beckner C and Whittinghill JR (1988) Comorbidity of mental disorders in the post-mortem diagnosis of completed suicide in children and adolescents. *Journal of Affective Disorders* **15**: 227–233.

Shneidman ES (1976) *Suicidology: Contemporary Developments*. New York: Grune and Stratton.

Shneidman ES (1994) The psychological autopsy. *American Psychologist* **49**(1):75–76.

Shneidman ES and Farberow NL (1961) Sample investigations of equivocal deaths. In: Farberow NL and Shneidman ES (eds) *The Cry For Help*. New York: McGraw-Hill.

Shneidman ES and Farberow NL (1970) Sample psychological autopsies. In: Shneidman ES, Farberow NL and Litman RE (eds) *The Psychology of Suicide*. New York: Science House.

Weisman AD and Kastenbaum R (1968) The psychological autopsy: a study of the terminal phase of life. *Community Mental Health Journal* (Monograph No. 4). New York: Behavioral Publications.

Zander M and Henderson P (1993) *Crown Court Study. The Royal Commission on Criminal Justice*. London: HMSO.

Appendix I

Outline for psychological autopsy (Shneidman, 1976)

1. Identifying information for victim (name, age, address, marital status, religious practices, occupation and other details)
2. Details of the death (including the cause or method and other pertinent details)
3. Brief outline of victim's history (siblings, marriage, medical illnesses, medical treatment, psychotherapy, previous suicide attempts)
4. 'Death history' of victim's family (suicides, cancer, other fatal illnesses, ages at death and other details)
5. Description of the personality and lifestyle of the victim
6. Victim's typical patterns of reaction to stress, emotional upsets and periods of disequilibrium.
7. Any recent (from last few days to last 12 months) upsets, pressures, tensions or anticipations of trouble
8. Role of alcohol and drugs in (1) overall lifestyle of victim, and (2) in his or her death
9. Nature of victim's interpersonal relationships (including with physicians)
10. Fantasies, dreams, thoughts, premonitions or fears of victim relating to death, accident or suicide
11. Changes in the victim before death (habits, hobbies, eating, sexual patterns and other life routines)
12. Information relating to the 'life side' of victim (upswings, successes, plans)
13. Assessment of intention, i.e. role of the victim in his or her own demise
14. Rating of lethality
15. Reactions of informants to victim's death
16. Comments, special features, etc.

Appendix II

Psychological autopsy guidelines (Ebert, 1987)

1. **Alcohol history**
   (a) Collect family history
   (b) Research amount ingested regularly
   (c) Research evidence of binge drinking
   (d) Research evidence of blackouts (known from friends, family, acquaintances)
   (e) Research evidence of driving under the influence of alcohol
   (f) Research evidence of alcohol-related offenses
   (g) Research evidence of family problems (alcohol-related)
   (h) Research evidence of work difficulties connected to alcohol
   (i) Research evidence of blood level ($g\,l^{-1}$) at time of death

2. **Suicide notes**
   (a) Examine content
   (b) Examine style
   (c) Have handwriting expert review writing style

3. **Writing**
   (a) Review any past writing by the deceased
   (b) Peruse any diary of the deceased
   (c) Examine school papers for topics of essays or term papers
   (d) Read letters to friends, family, coworkers, acquaintances

4. **Books**
   (a) Examine books of the deceased
      (i) Look for books on the occult, life after death, death
      (ii) Look for actual books on suicide
   (b) Assess books checked out of local libraries

5. **Relationship assessments**
   (a) Interview people who knew the deceased, including:
      (i) Close friends
      (ii) Close intimate heterosexual or homosexual companions
      (iii) Acquaintances
      (iv) Mother, father, siblings
      (v) Coworkers and supervisors
      (vi) Other relatives
      (vii) Physicians and/or mental health professionals
      (viii) Teachers
   (b) Construct level of intimacy on the basis of discussions with 'close' friends
   (c) Assess people's reactions to the victim's death
   (d) Secure a history of marriages and divorces
   (e) Examine relationship with children
   (f) Look for anger directed to particular people

6. **Marital relationship**
   (a) Note any significant problems that may have made the deceased person depressed
   (b) Look for history of extramarital relationships
   (c) Assess the overall quality of the relationship

7. **Mood**
   (a) Identify mood fluctuations
   (b) Look for symptoms of depression:
      (i) Weight loss
      (ii) References to depression
      (iii) Problems with memory
      (iv) Fatigue
      (v) Sleep disturbances
      (vi) Withdrawal
      (vii) Decreased libido
      (viii) Appetite and/or taste changes
      (ix) Constipation and diarrhea
   (c) Look for mood indicators during last few days:
      (i) Interview friends and family
      (ii) Interview anyone surrounding the deceased

8. **Psychosocial stressors (note and chart importance on Holmes & Rahe scale factors)**
   (a) Recent loss: deaths of people or pets
   (b) Relationship separations: divorce, breakup of significant relationships
   (c) Loss of job
   (d) Legal and financial problems
   (e) Demotion, promotion and so on
   (f) Reaction to stressors
   (g) Move to a new location

9. **Presuicidal behavior**
   (a) Giving away important possessions
   (b) Paying up insurance policies
   (c) Payment of debts
   (d) Arrangement for children and pets
   (e) Sudden order in deceased's life
   (f) Change or initial creation of a will

10. **Language**
   (a) Identify any specific references to suicide (deceased may have stated, 'Have a party in remembrance of me' or 'You won't have to worry about me anymore')
   (b) Note any changes in language before suicide

(c) Analyze language (tapes, recollections of conversations, writing) for morbid content

11. Drugs used

(a) Identify all drugs used by deceased
(b) Assess interactional effects of legal and illegal drugs in use

12. Medical history

(a) Review complete medical history
(b) Note any unusual symptoms or diagnoses
(c) Note any terminal illnesses or diagnoses

13. Reflective mental status examination of deceased's condition before death

(a) Orientation
(b) Memory
(c) Concentration
(d) Mood and affect
(e) Hallucinations or delusions
(f) Cognition, IQ
(g) Language
(h) Judgment

14. Psychological history

(a) Look for previous suicide attempts (type, method)
(b) Assess reason for treatment if involved in therapy
(c) Research evidence of depression, manic depression (bipolar disorder)
(d) Research past psychiatric hospitalizations
(e) Examine diagnoses
(f) Examine evidence of impulsive behavior
(g) Examine any recent or past psychological tests (e.g. was the victim given the Rorschach test and was the suicide constellation served via the Exner system?)

15. Laboratory studies

(a) Examine ballistics
(b) Evaluate powder burns on hands and body

16. Coroner's report

(a) Conduct complete drug screen
(b) Identify any poisons
(c) Read for detailed description of physical functioning/health of deceased at time of death

17. Motive assessment

(a) Make a chart divided four ways: murder, suicide, accident and natural, recording data to support each as it is uncovered
(b) Report the possible reasons for suicide
(c) Report the possible reasons why subject could have been murdered (identify enemies, illicit activities)

18. Reconstruction of events occurring on the day before deceased's death

(a) Make a step-by-step chart of subject's movements and activities
(b) Form a chronological history of the victim that immediately preceded death

19. Assess feelings regarding death as well as preoccupations and fantasies

20. Military history

(a) Look for evidence of difficulty adjusting such as letters of counseling (LOC), letters of reprimand (LOR), Article 15 action (A15), or court-martial proceedings. (Note: A15 is a form of non-judicial punishment for offenses not serious enough to warrant a court-martial and include repeated lateness, driving under the influence of alcohol, sleeping on duty or negligence on duty. Punishment from an A15 can include reduction in rank, fines or removal from duty.)
(b) Attempt to secure job ratings (airman promotion rating and officer effectiveness rating)
(c) Look for decorations or awards
(d) Notice whether deceased was in a combat zone at any time
(e) Look for evidence of posttraumatic stress disorder in Vietnam veterans
(f) Determine the number of assignments and which were at the request of the victim

21. Death history of family

(a) Examine history for suicide by other family members
(b) List immediate deceased family members and their mode of death

22. Family history

(a) Identify family members and relationships with deceased
(b) Examine the socioeconomic status of family
(c) Identify any conflicts that occurred before death of the victim

23. Employment history

(a) Identify number and types of jobs (high-risk work may indicate the existence of subintention behavior for quite some time)
(b) Look for repetitive problems
(c) Assess whether any problems existed before death (e.g. coworker conflict, failure to progress as planned)
(d) Note any disciplinary action

24. Educational history

(a) Assess educational level
(b) Identify any problems with teachers or subjects
(c) Note special interests or topics (e.g. in particular, look for special interests in death)

24. Familiarity with methods of death

(a) Examine belongings for guns, knives (e.g. the deceased may have had five or six loaded weapons around his or her house regularly)
(b) Look for lethal drugs
(c) Note deceased's interest in and knowledge of weapons

25. Police report

(a) Critical facts will be obtained by review of the police investigation
(b) Pay special attention to ballistics data

PSYCHOLOGY AND PSYCHIATRY

Contents
Overview
Pyschiatry
Psychology

Overview

C R Hollin, University of Leicester, Leicester, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0604

Introduction

This overview of forensic psychology and psychiatry first gives a distinction between the two disciplines, primarily in terms of professional training. Taking a historical perspective, the development of the two disciplines and their subject matter is then discussed in detail. This historical perspective is then used to set in context current practice within forensic psychology and psychiatry.

Psychology and Psychiatry

What are forensic psychology and forensic psychiatry? The basic answer is that forensic psychologists have initially qualified as psychologists, then trained at postgraduate level to practice in the forensic field. Forensic psychiatrists, on the other hand, have qualified in medicine, then taken further training to qualify as psychiatrists with a forensic speciality. In practice, forensic psychiatrists also hold statutory powers and responsibilities, such as prescription of medication and supervision of patients, not given to other professions. Thomas Grisso has suggested that there are several further distinctions that can be drawn between forensic psychology and psychiatry. First, psychiatrists are trained to consider abnormalities in an individual's biological and medical functioning; psychologists do not have this speciality, but are trained to focus on personality, behavior and broader social dimensions. Second, psychiatrists tend to draw their knowledge from clinical observation and research with clinical samples; psychologists are inclined to rely more heavily on experimental data. Third, psychiatrists lean towards assessment through clinical interview and observation; psychologists are more likely to use psychometric and other psychological tests.

Traditionally, forensic psychology and psychiatry have their roots in the application of psychological

knowledge and expertise to gather evidence for judicial purposes. It is undoubtably the case that this traditional use of the term 'forensic' is closest to the dictionary meaning. For example, *The Concise Oxford Dictionary* entry for 'forensic' gives, 'of, used in, courts of law'. In recent years the use of the term 'forensic', particularly by psychologists, has broadened considerably to encompass not just the courts, but any topic connected with crime and law. Thus, some texts on forensic psychology cover subjects such as the police, treatment of offenders, risk assessment and theories of criminal behavior. Now, as will be seen, there is a very strong case for etymological correctness in defining the boundaries of forensic psychology.

Ronald Blackburn, a leading British psychologist, has criticized this indiscriminate use of the term forensic to refer to any psychological activity 'vaguely connected with the law'. Blackburn's case is that unless it adheres to the correct, dictionary definition of the term 'forensic', then a profession that calls itself forensic psychology struggles to find an identity. Without a professional identity, we are left with the question of what are the unique skills and abilities that the forensic psychologist can offer? The root of the issue is that the term 'forensic psychology' has metamorphosed from a description of an *application* of psychology to describing a *kind* of psychology.

Discussion of forensic psychology and psychiatry here will be limited to the traditional sense and use of the term. In seeking to understand forensic practice, it is illuminating to look at the historical development of the two professions.

Historical Development

Forensic psychology

It is generally taken that forensic psychology dates from 1896, when Albert von Schrenk-Notzing appeared as an expert witness in a Munich court, offering testimony at the trial of a man accused of murdering three women. The specialist knowledge that von Schrenk-Notzing, a student of Wilhelm Wundt, the founder of the first psychological laboratory, was able to put before the court was based on experimental psychological evidence on the workings of memory. Specifically, von Schrenk-Notzing argued that the memory of the witnesses to the crime before the courts was unreliable because the witnesses had become confused between their own memories for the event and the pretrial publicity. While a string of great figures in the history of psychology, including James Cattell and Alfred Binet, made contributions to the emerging field of forensic psychology around the turn of the century, the greatest of the early luminaries was another of Wundt's pupils, the German psychologist Hugo Münsterberg (1863–1916).

In his book *On The Witness Stand*, published in 1908, Münsterberg argued that the emerging discipline of psychology could be beneficial to the process of justice. These benefits, Münsterberg claimed, lay not only with the application of psychological knowledge to the eyewitness, but extended to other matters such as the working of the jury. Thus, Münsterberg made the case for the psychologist's role in the courtroom as an expert witness to inform the process of justice on matters psychological. Münsterberg's arguments drew fierce attack from some quarters of the legal profession, anxious to resist the forays of psychologists into the courtroom.

In parts of the world other than Germany, psychologists also began to offer a contribution to courtroom proceedings. In America in 1921 a psychologist argued in court that a victim's testimony was doubtful because of low intellect. Similarly, in France Alfred Binet worked on the issue of the reliability of eyewitness memory.

After World War II, the profession of psychology grew in scope with the establishing of specialist educational and clinical psychologists. At first, these specialists provided psychological material for inclusion in medical court reports. However, as time passed, psychologists were permitted to submit court reports in their own right, rather than have their work subsumed into a report written by another profession. As the discipline of psychology progressed from the 1950s to the present day, so psychologists gave their attention to an increasing range of courtroom issues. Importantly, the role of the psychologist as expert witness developed considerably: several surveys, summarized in **Table 1**, showed that contemporary psychologists are testifying in court on a range of topics. However, psychologists were also cultivating a research agenda that would impact on two areas: first, research that would lead to greater levels of understanding with respect to the

Table 1 Typical subjects found in UK psychologists' court reports

Post-traumatic stress disorder
Compensation
Childcare
Suitability for treatment
Fitness to plead/stand trial
Reliability of witness statements
Reliability of confessional evidence
Diminished responsibility

Adapted from Gudjonsson (1996).

reliability of evidence; second, research that would increase dramatically the level of psychological insight into law and its application in the courtroom. Some of this research will be considered in detail in the entry on forensic psychology.

Forensic psychiatry

If the discipline of psychology has a short history, most often seen with its genesis in Wundt's laboratory, as noted above, quite the opposite is the case for forensic psychiatry. The medical profession, of which psychiatry is a branch and forensic psychiatry a speciality within that particular branch, has a long history. Indeed, the medical historian Roy Porter traces the evolution of medicine to the end of the last Ice Age, around 10 000–12 000 years ago, as the organization of societies shifted from hunter-gatherer to farming in order to sustain life. Porter suggests that the transmission of pathogens from animal to human led to an explosion in diseases such as tuberculosis, smallpox, measles and typhoid. As the population settled together in ever greater numbers, so the spread of parasites such as worms, flukes and lice increased. Similarly, changes in diet as agriculture became the main means of food production brought about conditions related to low protein intake and other dietary factors. Thus, population growth and changes in societal structure heralded the onset of epidemics such as the plague. As history unfolds, each evolutionary shift brings its benefits but also has a health cost. The industrial revolution brought prosperity (to some) as well as illness caused by pollution; the demands of modern-day living are associated with stress-related conditions such as hypertension and excessive use of drugs and alcohol.

As illness and disease took their toll in terms of human suffering, so people tried to find ways to heal the sick. With the advent of healing, medicine was born. The historical record shows that throughout antiquity civilizations developed their own understandings of the causes and cures of illness. As history unfolds, so important advances were made: we see, for example, improved understanding of human anatomy, appreciation of the complexity of disease, discovery of new drugs, and technological progress from microscopes and X-rays to brain scans and microsurgery.

It was long recognized that there were some people whose illness was of a special type; madness was as familiar to the ancient Greeks as it is in contemporary society. However, there was no special branch of medicine dedicated to relieving the suffering of the mad. The treatment and care of the mad was left to families and communities or, if all else failed, to madhouses run by charities or religious groups. It was not until the middle of the eighteenth century that psychiatry began to appear as a specialized branch of medicine. The rise in the use of asylums for the insane across much of Europe, and later America, heralded the increasing involvement of members of the medical profession with those suffering from mental disorder. The treatment of those people held in asylums was not always gentle: shock treatments, confinement, purging and physical restraint were common practices in the struggle to find a cure for insanity.

The great pioneers of psychiatry, including Benjamin Rush (1745–1813), Henry Maudsley (1835–1918) and Emil Kraepelin (1856–1926), set in train a change of thinking, emphasizing the need for medical understanding and treatment of the mentally disordered. The influence of such thought, coupled with social reaction against the conditions in the asylums and the discovery of psychoanalysis as a treatment for mental disturbance, culminated in the birth of the psychiatric hospital. So it came about that the treatment of the mentally ill became the province of the medical profession. As psychiatry found a professional identity, so psychiatrists became the experts on mental disorder.

In law, the insane have traditionally been accorded a special status, in that it is accepted that they should not be punished for criminal acts. The reasoning behind this is plain: the legal concept of *mens rea* holds that if a person is to be held responsible for his or her actions then it must be the case that in committing a crime they have acted of their own free will. Can it be said that the mentally ill are responsible for their actions, including criminal acts? Two particular legal cases are seen as critical in developing an answer to this question.

James Hadfield, soldier and member of the Duke of York's bodyguard, received head injuries from sabre wounds when fighting in the Battle of Lincelles, Flanders, in 1794. From accounts of the case, it is evident that Hadfield must have suffered brain damage, precipitating bouts of mental instability. It was said that during these bouts of insanity he would threaten the life of his child because God so commanded. Further, he proclaimed that he must die to save the world, yet he must not kill himself. Such statements would be taken now as clear evidence of delusions following head injury.

Hadfield devised a plan that would bring about his death: he would assassinate the monarch, King George III, for which the penalty would be execution. In 1800 he made his attempt, firing a pistol at the King as he entered a theatre. His shot missed and he was captured by bystanders, charged with high treason

and brought to trial. Hadfield's trial excited public interest and concern in the social context of both a growing awareness of the plight of the mentally ill, and widespread sympathy for Hadfield as a soldier wounded in battle.

During the trial, Hadfield's defense, delivered by Thomas Erskine, sought to refine the arguments regarding the legal position of the mentally ill. Erskine could not use the so-called 'wildbeast test' as Hadfield was clearly lucid and able to comprehend the court proceedings. Instead, Erskine used a more subtle approach, advancing the argument that Hadfield's delusions at the time of the offense must be taken into consideration. If the court agreed that Hadfield was deluded at the time of the assassination attempt then, Erskine argued, he cannot be considered guilty of a crime. Medical and lay evidence was called to support the defense's position regarding Hadfield's madness. Erskine's strategy worked and the Lord Chief Justice halted the trial, directing the jury to find Hadfield not guilty, 'he being under the influence of Insanity at the time the act was committed'.

This judgment had two important consequences: first, it created a new class of offender, then called 'criminal lunatics'; second, it created a practical problem – if Hadfield was not a criminal then it follows that he could not be sent to jail. However, Hadfield was judged to be a danger to the public and therefore some form of secure detention was required. As will be seen presently, the solution to this dilemma was an important one in the development of forensic psychiatry. However, turning to the second important case, the name of Daniel McNaughton (spelt M'Naghten in some texts) rises to attention.

From case records dating back to the 1840s, it is highly likely that McNaughton was suffering from paranoid schizophrenia. Specifically, McNaughton believed that he was being persecuted by Catholic priests and by the Tories. This delusional belief led him to the decision that to end his persecution he would kill the Tory Prime Minister, Sir Robert Peel. In 1843, McNaughton made his attempt on the life of the Prime Minister but, through mistaken identity, shot and killed the Prime Minister's private secretary. McNaughton was arrested at the scene of the crime and sent for trial. The weight of eminent medical opinion at the trial was that McNaughton was insane and the jury returned a verdict of not guilty on grounds of insanity. However, McNaughton's insanity was not of a florid type, leading to obvious loss of control: it was clear from his own statements that he knew very well what he was doing at the time of the crime and acted deliberately and was fully aware of his actions. Despite McNaughton's plea that he was driven by his persecutions, it was the calculated quality of the seemingly cold-blooded killing that brought about a wave of public protest after the trial. Accordingly, leading judges of the day were asked to inform the Government on the interpretation of the laws of the land with regard to the legal position for cases of insanity and crime. In response, the judges formulated the McNaughton test, or McNaughton rules as they are sometimes called.

The McNaughton test, given several caveats, elucidates several key principles. First, everyone who stands before a court is to be presumed sane and responsible unless proven otherwise. Second, an insane person is punishable if he or she knew at the time of the offense that the act was wrong. Third, to establish an insanity defence it must be proved that at the time of the offense the accused did not know the nature of the act; or if he or she were aware of their actions, then they did not know that their actions were wrong. Thus, to establish an insanity defense, the test is: does the accused know the nature of the act? If they do know, do they know it is wrong? These are powerful questions demanding fine skills and knowledge to provide answers: some members of the medical profession of the day began to specialize in understanding the criminal lunatic so as to inform the courts on just such matters.

As the courts began to grapple with the implications of cases such as Hadfield and McNaughton, the issues raised were not just legal but were also practical in nature. If such people were not criminals, could they be sent to jail? If they were judged to be dangerous, could they be released into the community? The judge in the Hadfield case did send Hadfield to jail but expressed doubts about his authority to do so under the law of land. A series of parliamentary acts from 1800 onwards therefore began to put in place legislation to allow criminal lunatics to be admitted to asylums for the mentally ill. Hadfield was duly taken from jail and admitted to Bethlem Hospital. Hadfield was not a model patient: he killed another patient and escaped from the hospital, spending time at large before his recapture and return to Newgate Gaol. Deep concern was expressed that asylums were being asked to admit patients for whom the facilities, personnel and buildings were not suited.

The first response was the development of hospital 'criminal wings': these wings were secure, with locks, bolts and bars, but staffed by doctors and nurses to treat those detained. In 1816, Hadfield was sent to the criminal wing at the Bethlem Hospital where he remained until his death in 1849, when he was 70 years of age. While the criminal wings sufficed for a time, by the late 1850s concern was once again being expressed about the security, safety and standards on these wards. In 1860, permission was given for the

building of Broadmoor Criminal Lunatic Asylum in Berkshire; Broadmoor opened in 1863 and began receiving its first patients, decanted from the criminal wings of hospitals, to fill its 500 beds.

Broadmoor proved to be the first of a series of secure institutions for the criminal lunatic: Rampton followed in 1914, then Moss Side in 1933 and finally Park Lane in 1974 (the latter two being amalgamated to form Ashworth); Carstairs in Scotland performs a similar function. More recent legislation has given these institutions the generic title of 'special hospital', and they are charged with detaining mentally disordered offenders under conditions of maximum security. Moving to much more recent times, a new wave of services has developed, including the Regional Secure Units, which offer treatment for mentally disordered offenders in settings of medium security, and in the community.

As services have expanded, so the importance of understanding the criminal lunatic has increased accordingly and forensic psychiatry has rapidly developed. Thus, legal thinking has produced the classifications of the mentally disordered offender in operation today; academic study of the relationship between mental disorder and criminal behavior has become increasingly sophisticated; decisions on competency to stand trial and risk assessment are more complex; and the design, running and evaluation of services for the mentally disordered offender engage the attention of many profrssionals. It is clear that these developments, while not identical, have followed a similar line in North America and many countries in Europe.

Current Practice

It would be wrong to draw an absolute line of demarcation between the practice of forensic psychologists and psychiatrists. The issues highlighted below are predominately but not exclusively the concern of the respective professions.

Forensic psychology

Taking a strict definition of forensic psychology, what might a forensic psychologist offer the court with respect to the evidence placed before it? There are two strands to follow in this regard. First, the psychologist can present the findings of psychological research on different types of evidence; second, the psychologist may offer an expert opinion in a given case. In terms of the first point, research has focused on both eyewitness evidence and confessional evidence. With respect to the second point, Lionel Haward, the most eminent British forensic psychologist, has described four roles that the forensic psychologist might take as an expert witness.

Actuarial role In this role the forensic psychologist presents evidence regarding the probability of an event, drawing on extant statistical information. For example, the psychologist might provide actuarial evidence regarding the sequelae of some psychological trauma and its impact on social functioning.

Advisory role The psychologist may be asked to provide an opinion on another expert's report, advising on the accuracy and robustness of the report and which issues bear questioning when the report's author is under examination. The advice may be given before the trial or during the trial as the other expert's testimony unfolds.

Clinical/assessment role In this role the forensic psychologist is concerned to present the court with evidence regarding state of mind. This may relate to the presence of psychological disorder or dysfunction, as, for example, whether or not an individual manifests post-traumatic stress disorder. However, Haward extends this role to include not only the assessment of disorder or dysfunction but also the assessment of normal human functioning. For example, Gudjonsson and Haward note a case in which it was important to establish that the accused was left-handed: a series of observational, neurological and psychological tests were used to present evidence to the court regarding this issue.

Experimental role The credibility of evidence may hinge on whether human performance under given conditions is likely to be accurate. Could a witness have really seen a face under low levels of light? Is it possible to make accurate judgments of another car's speed when travelling in a moving vehicle? Such questions are open to experimental investigation by simulating the conditions at the time of the incident and testing performance. Such investigations allow the psychologist to give the court an empirically informed probability of the reliability of evidence in a given case.

Forensic psychiatry

The role in court of the forensic psychiatrist may be identical to those defined by Haward for the forensic psychologist, although reflecting a different professional perspective. Thus, forensic psychiatrists might draw on the research on the relationship between mental disorder and criminal behavior. Alternatively, in a given case a forensic psychiatrist might be called

on to give an expert opinion with respect to fitness to stand trial, or risk assessment, or treatability.

Concluding Comment

While there are differences between forensic psychology and psychiatry, it would be wrong to think that these differences must lead to adversity. In the labyrinth of forensic practice it would be a mistake for any profession to believe it had a monopoly on the truth. It is only when forensic psychologists and psychiatrists work together (and with other professions such as forensic nurses, probation officers and the police) that the true potential of each discipline can be achieved.

See also: **Expert Witness**: Qualifications and Testimony. **Psychology and Psychiatry**: Psychiatry; Psychology. **Investigative Psychology**.

Further Reading

Bartol CR and Bartol AM (1994) *Psychology and Law: Research and Application*. Pacific Grove, CA: Brooks/Cole.

Blackburn R (1996) What is forensic psychology? *Legal and Criminological Psychology* **1**:3–16.

Bluglass B and Bowden P (eds) (1990) *Principles and Practice of Forensic Psychiatry*. Edinburgh: Churchill Livingstone.

Bull R and Carson D (eds) (1995) *Handbook of Psychology in Legal Contexts*. Chichester: Wiley.

Forshaw D and Rollin H (1990) The history of forensic psychiatry in England. In: Bluglass B and Bowden P (eds) *Principles and Practice of Forensic Psychiatry*, pp. 61–101. Edinburgh: Churchill Livingstone.

Grisso T (1993) The differences between forensic psychiatry and forensic psychology. *Bulletin of the American Academy of Psychiatry and Law* **21**:133–145.

Gudjonsson GH (1996) Psychological evidence in court: results from the 1995 survey. *Psychologist* **5**:213–217.

Gudjonsson GH and Haward LRC (1998) *Forensic Psychology: A Guide to Practice*. London: Routledge.

Gunn J and Taylor PJ (1993) *Forensic Psychiatry: Clinical, Legal and Ethical Issues*. Oxford: Butterworth-Heinemann.

Haward LRC (1981) *Forensic Psychology*. London: Batsford.

Hollin CR (1989) *Psychology and Crime: An Introduction to Criminological Psychology*. London: Routledge.

Kagehiro DK and Laufer WS (1992) *Handbook of Psychology and Law*. New York: Springer.

Kapardis A (1997) *Psychology and Law: A Critical Introduction*. Cambridge: Cambridge University Press.

Porter R (1997) *The Greatest Benefit to Mankind: A Medical History of Humanity from Antiquity to the Present*. London: HarperCollins.

Prins H (1995) *Offenders, Deviants or Patients?* 2nd edn. London: Routledge.

Psychiatry

C R Hollin, University of Leicester, Leicester, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0605

Introduction

As is evident from the major texts, the subject matter of forensic psychiatry lies in the application of psychiatric knowledge to offender populations. In principle, this brief extends to all offender groups, although in practice it is the juxtaposition between mental disorder and criminal behavior that is the traditional concern of forensic psychiatry. With respect to mental disorder and criminal behavior, three populations can be defined: (1) those people with a mental disorder who do not commit crimes; (2) offenders who are not mentally disordered (which includes the majority of offenders); (3) those who display both mental disorder and criminal behavior. It is this third group, mentally disordered offenders, that is of particular concern to forensic psychiatry and so forms the main interest of this article.

The starting point in any discussion of mental disorder and crime lies in the definition of mental disorder. In general psychiatric practice, mental disorder is diagnosed by reference to formal diagnostic systems, such as the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders*, the revised fourth edition (DSM-IV-R) of which is in current usage. For the purposes of law, however, a legal definition is required: for example, the Mental Health Act, 1983 distinguished four manifestations of mental disorder: mental illness; arrested or incomplete development of the mind; psychopathic disorder; and any other disorder or disability of mind. There is not an exact correspondence between formal psychiatric diagnostic systems and legal definitions, so that in practice forensic psychiatrists must work with the courts in deciding upon issues of mental disorder as the need arises.

The overlap between mental disorder and criminal behavior has concerned criminologists, lawyers and psychiatrists for many years. There are several reasons for this long-standing concern. It is intrinsic to the dignity and fairness of the legal system that it does not penalize those members of society who are mentally unfit. Any individual who cannot function mentally is seen as potentially unreliable and therefore cannot be assumed to be credible; this failing may weaken and call into question the fairness of the process of justice. Similarly, an accused who is not fully competent cannot reliably be said to be making

informed decisions with respect to their case. Finally, the imposition of criminal sanctions, such as imprisonment, is not held to be fair and just if the accused is not fully responsible for their actions.

From the wide range of concerns of forensic psychiatry, three areas have been selected for discussion here, as they encapsulate issues at different legal stages with regard to the mentally disordered offender. These three topics are: assessment of competence prior to committal; risk assessment with respect to placement; and treatment during placement.

Assessment of Competence

It is a fundamental aspect of law that those being tried are seen to be responsible for their actions: first, the accused is responsible for their actions at the time that the crime was committed; second, that the accused is fit to stand trial. Each of these will be considered in turn.

For a crime to be committed in law there must be two elements, *mens rea* and *actus reus*. While both terms defy simple definition, the former may be thought of as the intention to commit an act while knowing it was wrong; the latter as the act (or failure to act, as in negligence) itself. Thus, the act may be committed but unless there can be shown to be guilty intent then there is no case to answer. The obvious example of this principle lies in the notion of an age of criminal responsibility: a child may commit a serious act but not be held responsible legally for their actions and therefore cannot be held to have committed a crime. Since the trial of Daniel McNaughton (M'Naghten), the law has evolved to take account of the view that the 'mad' or 'insane' or mentally disordered cannot be held responsible for their actions. In other words, those not seen to be responsible for their actions can be said to have *actus reus* but not *mens rea*.

The practical difficulty in the case of the mentally disordered offender lies in knowing whether the accused was aware of their actions and, if so, whether they knew that their act was wrong *at the time that they committed the act*. In cases of mental disorder it falls to the defense to mount an 'insanity defense'; that is, to plead that the accused is, in fact, not guilty by reason of insanity at the time the act was committed. There are variations on this theme across different legal systems; however, the task of the forensic psychiatrist remains the same in such cases. That is, the psychiatrist must look back ('postdict') to give a professional opinion of the accused's state of mind at the time of the crime. Given that there may be months or even years between the act and the need for a 'psychiatric postdiction', it is clear that such assessments can never be an exact science.

While the moral and legal issues are similar with respect to competency to stand trial (variously referred to as 'unfit to plead' or 'under disability'), the practical task facing the forensic practitioner is placed in the present rather than the past. It therefore falls to the forensic psychiatrist to inform the court whether the accused is mentally disordered, within the legal meaning of the term, such as to impair understanding of the charges faced. There are several protocols that have been developed to aid assessment of competency, such as Competency Screening Test and the Georgia Court Competency Test. It is unlikely, however, that these instruments will replace the traditional clinical skills of the forensic psychiatrist.

Risk Assessment

What is risk assessment? To conduct a risk assessment is to undertake an exercise in prediction: risk assessment is an estimation of the likelihood that some future event will occur. The estimation of risk is a skill that is required in many areas of life: financiers estimate the risks of their investments; insurers estimate the risks of fire, theft, car accidents; and surgeons estimate the risks of operations.

In forensic psychiatric practice with mentally disordered offenders the assessment of risk is a prime concern. It falls to forensic practitioners to make decisions regarding admittance to secure psychiatric facilities, and similarly to decide on recommendations of discharge from security and return to the community. Thus, it falls to practitioners to estimate the risk of an individual inflicting more harm on the community. In essence, practitioners are faced with a classic decision-outcome matrix as shown in **Fig. 1**.

Consider the legal decision to send a violent mentally disordered person to a maximum security hospital. It is the role of practitioners to inform the court of their prediction as to whether the level of risk of that person being a continued danger to the public (a key criterion) is such as to justify the decision of detaining that individual for treatment in conditions

	Actual Yes	Actual No
Predict Yes	True positive (Hit)	False Positive (Miss)
Predict No	False negative (Miss)	True negative (Hit)

Figure 1 Hits and Misses in prediction.

of security. Suppose the psychiatrist predicts Yes and, indeed, if the individual had stayed in the community there would have been more victims (actual Yes) then a decision to admit to hospital is correct (a Hit) and the public will be protected. However, suppose the prediction is Yes but in fact the individual would not have been a continued danger to the public (actual No). Now the person will wrongly be admitted to hospital (a Miss), with all the attendant issues with respect to human rights and public spending.

Alternatively, the practitioner might predict No and the person is not detained but then commits further offences in the community (actual Yes). In this event (a Miss), of which there have been several cases widely reported in the popular media, the scene is set for a public inquiry. Finally, the prediction may be No and the individual remains in the community and does not reoffend (actual No); obviously, this is a good decision (a Hit).

If the same logic is applied to decisions about release from secure hospital conditions back into the community, then it is plain that in its assessment of risk forensic psychiatry is charged with a heavy responsibility. In this light, the topic of risk assessment is of fundamental importance.

Terminology

A term commonly used in risk assessment is the 'criterion of risk'. This term refers to an exact description of the event that is being assessed: in other words, a clear definition of the concern of a risk assessment. For example, in physical medicine the criterion of risks might be the occurrence of heart disease or cancer, or even particular forms of illness such as angina or breast cancer.

Once the criterion of risk is clear, then to estimate risk, the associated predictors (or risk factors) must be identified. Thus, for the criterion of heart disease, the predictors might include a family history of heart disease, a history of smoking, a pattern of unhealthy eating and lack of exercise, a high cholesterol level, high blood pressure and a stressful occupation. It is important to emphasize that the presence of these predictors does not mean that the realization of the risk is inevitable. Rather, the presence of a predictor increases the likelihood that the event will occur, recognizing that some predictors carry more weight than others. Further, it is generally the case that the greater the number of extant predictors in a given case, the greater the risk becomes. Thus, an individual with a family history of heart disease, who smokes 20 cigarettes a day and has high blood pressure is more likely to suffer a heart attack than another individual who has a moderately high cholesterol reading. The important word here is *likely*: for neither individual is a heart attack certain, but the risk would certainly be seen as higher in the former compared with the latter.

The identification of predictors of risk is not a simple task. The issue facing practitioners is how to know exactly what aspects of an individual's history and current functioning should be of concern in order to estimate a given risk. The research literature traditionally makes the distinction between static predictors of risk and dynamic predictors of risk.

Static risk predictors are historical or background factors, the presence of which is known to elevate the risk of a given criterion. In the example given above, a family history of heart disease would be a static risk predictor. Again, the presence of static predictors, which by definition cannot change, do not mean that the event is predetermined: the presence of static predictors simply raises the probability of the event happening in the future.

Dynamic risk factors are aspects of an individual's current functioning; thus, smoking, an unhealthy diet and a stressful job are all dynamic risk factors in relation to heart disease. In practice, it is often a combination of static and dynamic predictors that gives the strongest basis by which to predict risk.

In summary, when completing a risk assessment system it is crucial to have clearly defined risk criteria. Once the risk criteria are set, the appropriate risk predictors can then be identified and methods of assessment put into place.

Methods

The classic distinction with respect to method of risk assessment is to be found in clinical and actuarial prediction of risk.

Clinical prediction The clinical method of risk prediction is based on professional judgment and decision-making. This judgment of risk may be made by an individual practitioner, or by a clinical team, or by a case conference.

Actuarial prediction The actuarial method of risk prediction relies on the use of statistical methods to identify the risk factors for a given behavior, say violent conduct. These statistically identified risk factors may then be combined into an algorithm to give a standardized risk assessment.

In developing systems of actuarial risk assessment, the first task is to identify the appropriate risk predictors for the given criterion. Initially, through the use of clinical records, case material and the research literature, potential risk factors are identified for the criterion of concern. An appropriate cohort is identi-

fied and measures are taken of all the potential risk factors. The cohort is then followed over a lengthy period (typically years) to establish which potential risk factors actually best predict the occurrence of the criterion of risk.

A good example of this approach to risk assessment is seen in research, conducted in Canada, concerned with the development of an actuarial prediction scale for violent recidivism in mentally disordered offenders. The cohort in this study consisted of more than 600 males treated in a maximum security psychiatric hospital. After a mean time at risk of almost 7 years, the cohort could be divided into those who had shown violent recidivism and those who had not shown this behavior. These two groups could then be compared across a range of demographic, clinical and offense-related measures, with an aim of determining statistically what particular factors differentiated the two groups and hence predicted violent recidivism.

As shown in **Table 1**, a total of 12 predictors of violent recidivism were identified using actuarial, statistical methods. These factors are efficient in the sense that they maximize the likelihood of making a Hit and minimize the likelihood of a Miss.

In order to make this research available for clinical practice, the Violence Prediction Scheme, for predicting violence in high-risk men, was produced. The Violence Prediction Scheme presents practitioners with full details on assessing the various predictors. Simply, in making an assessment each risk factor is assigned a score, the scores are totaled and the higher the score the greater the risk of violent recidivism.

Of course, the same procedure can be followed for other criteria. For example, sex offending in mentally disordered populations is an obvious concern: actuarial studies have identified several predictors of sexual recidivism. Another Canadian study reported that rapists presented a greater risk of recidivism than sex offenders against children. As shown in **Table 2**, a range of predictors of sexual offending was identified; however, while predicting sexual reoffending, the predictors did not discriminate between rape and offending against children.

There is now a great deal of information available to inform actuarial risk assessment with mentally disordered offenders but there remains much to be understood; for example, the research informing the development of the Violence Prediction Scheme reported that a DSM-III diagnosis of schizophrenia had a *negative* relationship with violent reoffending. It is, however, the case that there is a minority of people with schizophrenia who are at a greater risk of reoffending than others with the same diagnosis. The nature of the relationship between the finer aspects of the diagnostic symptoms of schizophrenia and offending *within* the population of people diagnosed as schizophrenic remains uncertain.

There is a long-running debate regarding the relative efficacy of clinical and actuarial prediction. Proponents of actuarial methods proclaim the superior accuracy and scientific basis of statistical prediction; proponents of the clinical method argue for the skill of the clinician in making individually based risk assessments and against extrapolating from broad-based research findings to the individual patient. Rather than one method being compared with the other, it would be advantageous to see controlled studies of a combination of methods.

Treatment and the Mentally Disordered Offender

Alongside court work, the delivery of treatment to the mentally disordered is a mainstay of forensic psychiatry services. Such services can be delivered in the community, in hospitals (both open and secure) and in prisons. The task of those offering treatment to the mentally disordered offender is complex. Alongside an awareness of risk to the public, and in some cases the need to work within conditions of security, forensic practitioners are faced with two tasks: the

Table 1 Actuarial predictors of violent recidivism

1. Level of victim injury in index offence
2. Never married
3. Female victim – index offence[a]
4. Failure on prior conditional release
5. Property offence history
6. Age at index offence
7. Alcohol abuse history
8. High score on psychopathy checklist
9. DSM-III personality disorder
10. Separation from parents under age 16
11. Victim injury in index offence[a]
12. DSM-III schizophrenia

[a] These factors are contraindicative of recidivism; i.e. their presence suggests that violent recidivism is less likely.
From Webster *et al*. (1994)

Table 2 Actuarial predictors of sex offence recidivism

1. Never married
2. Violent convictions
3. Admissions to corrections
4. Previous sexual convictions
5. Previous female victims
6. Previous male child victims
7. Deviant sexual interest
8. High score on psychopathy checklist

From Quinsey *et al*. (1995).

treatment of the disorder and lowering the risk offending. It is not always the case that the two are synonymous.

Take, for example, the case of a person with schizophrenia who exhibits angry outbursts and commits violent assault. The forensic practitioner is faced with the need to treat the schizophrenia, to treat the anger and to lower the risk of violent behaviour. It cannot be assumed that any or all of these three dimensions of the patient's behavior, i.e. the schizophrenia, anger and violent conduct, are related. The key clinical skill lies in conducting an assessment that will distinguish between treatment targets related to clinical outcome (i.e. improved mental health) and targets related to lowering the risk of harm to the public. How might the identification of risk factors work in practice? The example below highlights issues in the assessment and management of the risk of violence.

Risk assessment and management

As discussed above, research points to a range of actuarial predictors of violent behavior. In a comprehensive risk assessment system, a patient's files and case material would be systematically trawled for the presence of these predictors. Clinical interview and observation would then look for the presence of these and other dynamic predictors of violence (e.g. high levels of anger arousal). Thus, a combination of actuarial and clinical risk assessment of the violent conduct will alert clinical teams to the factors that are relevant for an individual patient. This assessment will, in turn, help the clinical teams in setting initial treatment plans to bring about change on the factors related to violence. The process of risk assessment will continue throughout a patient's contact with clinical services. It is likely that new risk factors will arise as clinical teams develop a fuller understanding of their patient. Finally, practitioners must be confident that their assessment and treatment has had the necessary effects, so that the individual can be released to the community.

When stated in this way the process might seem quite straightforward; in fact, it is far from being an exact science. The complexities are perhaps best illustrated by what is known of the effectiveness of treatment for those sent for treatment as psychopaths.

Treatment of psychopathy

Psychopathy is a legal term which probably best translates into clinical terminology as a personality disorder, or, more exactly, antisocial personality disorder. The characteristics of such individuals, who have generally committed a string of violent and sexually violent offences, is generally assessed in terms of their lack of remorse for their actions, lack of empathy, and deceit and manipulation of other people. The problems of definition and diagnosis aside, there is a (probably small) group of psychopathic/personality disordered people who are dangerous to society at large. Such individuals will be received into forensic services: what is the optimum way to treat them?

Following an extensive review of studies of the treatment of offenders with personality disorder, it was noted that a variety of therapeutic approaches have been tried: this includes psychopharmacological treatment, psychodynamic psychotherapy, and cognitive-behavior therapy. In concluding the review, the author comments that no particular approach has consistently been shown to be beneficial, but procedures that structure the therapeutic environment, such as the therapeutic community, can claim some positive effects.

However, to add to the confusion, those treatment studies that have focused on assessment of psychopathy (using a measure called the Psychopathy Checklist), rather than clinical diagnosis of personality disorder, suggest that such individuals show lower levels of motivation to participate in treatment. Further, they show significantly lower levels of clinical improvement, and are more likely to drop out of treatment before it is completed. Indeed, one outcome study went so far as to suggest that psychopathic offenders showed *higher* levels of recidivism after time in a therapeutic community.

Conclusion

While relatively few in number, mentally disordered offenders are of great public concern and set many challenges for forensic psychiatry. The complexities of the moral and legal issues with respect to responsibility, the intricacies of assessment and the problems of treatment are all live issues. In looking at the problems it is perhaps easy to forget that forensic psychiatry, as with other forensic specialities, has a short history. If the current pace of growth in knowledge can be maintained, answers to many of the questions set here will surely be found.

See also: **Legal Aspects of Forensic Science**. **Psychology and Psychiatry**: Overview.

Further Reading

Bartol CR and Bartol AM (1994) *Criminal Behavior: A Psychosocial Approach*, 4th edn. Pacific Grove CA: Brooks/Cole.

Bluglass B and Bowden P (1990) *Principles and Practice of*

Forensic Psychiatry. Edinburgh: Churchill Livingstone.
Dolan B and Coid J (1993) *Psychopathic and Antisocial Personality Disorders: Treatment and Research Issues*. London: Gaskell.
Grove WM and Meehl PE (1996) Comparative efficiency of informal (subjective impressionistic) and formal (mechanical, algorithmic) prediction procedures: The Clinical–Statistical Controversy. *Psychology, Public Policy, and Law* **2**:293–323.
Hare RD (1991) *The Psychopathy Checklist – Revised*. Toronto: Multi-Health Systems.
Harris GT, Rice ME and Quinsey VL (1993) Violent recidivism of mentally disordered offenders: the development of a statistical prediction instrument. *Criminal Justice and Behavior* **20**:315–335.
Howells K and Hollin C (1993) *Clinical Approaches to the Mentally Disordered Offender*. Chichester: Wiley.
Lykken DT (1995) *The Antisocial Personalities*. Hillsdale, NJ: Erlbaum.
Monahan J and Steadman HJ (eds) (1994) *Violence and Mental Disorder: Developments in Risk Assessment*. Chicago, IL: University of Chicago Press.
Prins H (1995) *Offenders, Deviants or Patients?* 2nd edn. London: Routledge.
Quinsey VL, Rice ME and Harris GT (1995) Actuarial prediction of sexual recidivism. *Journal of Interpersonal Violence* **10**:85–105.
Webster CD, Harris GT, Rice ME, Cormier C and Quinsey VL (1994) *The Violence Prediction Scheme: Assessing Dangerousness in High Risk Men*. Toronto: Centre for Criminology, University of Toronto.

Psychology

C R Hollin, University of Leicester, Leicester, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0606

Introduction

The German psychologist Hugo Münsterberg (1863–1916) set the agenda for future generations of forensic psychologists in nominating evidence and the workings of the courtroom among the leading concerns of forensic psychology. This article will give a flavor of the contribution of psychological research to these two important areas of forensic psychology.

The Psychology of Evidence

Eyewitness testimony

The work of psychologists with respect to the accuracy of eyewitness testimony provides an excellent example of the interplay between psychological theory, applied psychological research and real-life decisions in the courtroom. This area of work is important, given that is now clear that the fallibility of human memory lies at the heart of a catalog of wrongful convictions based on the testimony of eyewitnesses. The work of psychologists, such as Whipple and Münsterberg, at the beginning of the twentieth century was concerned to establish the reliability of eyewitness evidence as an issue for law and psychology alike. Later research, from the 1970s onwards, has sought to offer an understanding of eyewitness testimony within more sophisticated psychological models of human memory.

One model of the functioning of human memory suggests that there are three stages to the process of memory: (1) the acquisition of information about the world (i.e. perception); (2) retention (i.e. storing information in memory); and (3) retrieval (i.e. remembering). If this three-stage model is matched against the process of witnessing an incident, then the stages can be defined as follows: (1) witnessing the incident (acquisition); (2) waiting to give evidence (retention); (3) giving evidence (retrieval). Further, within these three stages there are different kinds of variables to consider; these variables can be classified as social, i.e. to do with interactions between people; situational, i.e. concerning the setting in which the event occurred; individual, as with the characteristics of the individual witness; and interrogational, i.e. regarding the process of eliciting testimony. As can be seen from **Table 1**, a range of factors has been shown to be related to the accuracy of eyewitness memory. Clearly it is beyond the scope of this article to discuss these in detail – the topic of eyewitness memory has generated a large body of evidence, summarized in a number of texts. A discussion here of the key issues will give an understanding of the field.

Acquisition: nature of the witnessed incident The effect on memory of the nature of the witnessed incident has been considered in several studies. In a typical study, observers are shown one of two short filmed incidents; both films involve the same people, but in one film the content will be violent in nature, perhaps including scenes of physical assault. Analysis of the participants' recall of the filmed events typically reveals that memory is significantly better for the nonviolent rather than the violent incident. An

Table 1 Variables in eyewitness memory research

Social	*Situational*	*Individual*	*Interrogational*
Attitudes	Complexity of event	Age	Artists' sketches
Conformity	Duration of event	Alcohol use	Computer images
Prejudice	Illumination	Cognitive style	Identity parades
Status of interrogator	Time delay	Personality	Mugshots
Stereotyping	Type of crime	Race	Instructions
	Target factors	Sex	Photofit
	Weapon		Training

excellent example is provided by a study in which separate groups of observers saw different versions of a filmed bank robbery: in one scene shots were fired but no one was hit; in the other a small boy was hit and was seen bleeding from a head wound. Those witnesses who saw the violent scenes gave less accurate recollection of the incident.

From this and other experimental evidence, can it be concluded that serious crimes lead to poorer memory performance, perhaps particularly so when a weapon is involved? In fact, it seems likely that it is the degree of violence that is important: less threatening events that engage the witness's attention in a stimulating manner will produce satisfactory memories of the event. The research evidence suggests that witnessing extremely violent and personally threatening crimes may well be associated with poorer performance.

Retention: discussion between witnesses If there is more than one witness to an incident, should witnesses be allowed to discuss events before giving testimony? The research evidence suggests that group discussion can lead to a more complete description of an event, but at the price of introducing a source of error. While having the potential to prompt memory, discussion can bring about errors of commission – that is, testimony that includes errors related to items that were not seen in the original incident. Further, group dynamics can influence witnesses to change their testimony to agree with other members of the group. On balance, it is a moot point whether discussion between witnesses is desirable in the interests of producing accurate testimony.

Retrieval: leading questions A body of research, perhaps most notably by the American psychologist Elizabeth Loftus, has shown that even the most subtle changes in question wording can influence witness testimony. In one study, witnesses to a filmed automobile accident were asked to give estimates of the speed of the cars when 'they ... into each other'. Different groups of witnesses were asked this question, with the verb being one of 'contacted', 'hit', 'bumped' or 'smashed'. The witnesses' estimates of the speed increased, following the order given above, from 31.8 miles (51.2 km) per hour to 40.8 miles (65.6 km) per hour. Further, when later requestioned, the witnesses who had been asked about the 'smash' were more likely to say in error that there was broken glass at the scene of the accident.

Witnesses can also incorporate misleading information into their testimony: in another study, observers were asked to judge the speed of the car 'as it passed the barn', although no barn had been seen. One week later, the observers were asked a set of questions about the film, including whether they had seen a barn; more than 17% of the 'primed' observers said that they had seen the fictitious barn.

It appears that misleading information is most likely to have an influence on testimony if the source of the information is seen to be of high status, or if the witness thinks the police have special knowledge about the incident, or if the misleading information relates to peripheral detail rather than central events.

Witness factors Witness factors, such as intelligence, personality, anxiety, age and sex, are active across all three stages of memory. Reviews of the research conclude that there is little evidence of such factors having any consistent effect on eyewitness memory. However, there are two exceptions to this general conclusion: the first is the research evidence on the relationship between witness confidence and accuracy; the second is the now substantial body of work on children as witnesses. The research evidence on the former suggests that it is prudent to be cautious about witness confidence: confidence in the accuracy of one's memory is not an index of accuracy. With respect to children as witnesses, there are grounds to argue that, given sensitive questioning, even young children can be reliable witnesses, even to traumatic events.

Confession evidence

For a person to confess to a crime is, surely, an admission of guilt beyond dispute. While people unconnected with a case do come forward and make false confessions, police investigation often shows beyond doubt that such admissions are indeed fake. Such confessions are sometimes given by people who are seeking attention or who are psychologically distressed in some way. On other occasions, suspects will knowingly give a false confession during police interrogation; they may do this to relieve themselves from the stress of interrogation or to protect another person. Such false confessions are given in the (sometimes mistaken) belief that the police investigation will inevitably show that their confession is untrue.

In the examples given thus far, the individuals, unless severely disturbed, are aware that their confession is false. Is it possible that a suspect could *wrongly* confess to a crime he or she did not commit?

The phenomenon of the 'coerced-internalized false confession' is seen when the suspect's own private version of the truth actually changes to come into agreement with the interrogator's version. How can such a transformation of memory take place? In order to build an understanding of the coerced-internalized false confession it is important to look at the characteristics of both the interrogation and the suspect.

The work of the forensic psychologist Gisli Gudjonsson has played an enormous role in helping to understand the processes that act to produce false confessions. Gudjonsson suggests that the suspect under interrogation is in fact taking part in a highly unusual 'closed' social interaction: interrogation occurs in a closed room; the suspect is closed off from his or her normal surroundings; and the suspect's attention is closed in on the interrogator. Further, the interrogator holds the balance of power and therefore controls all aspects of the interaction. Trained in interrogational techniques based in the psychology of persuasion, the interrogator's focus is on gaining a confession from the suspect.

Gudjonsson suggests that some suspects will cope with the stress of interrogation by actively dealing with the situation and resisting the pressure to confess. However, other suspects will become passive and take a helpless stance, seeking to avoid confrontation with the interrogator, so reducing their emotional stress. Suspects who react in this passive way may be overly suggestible with regard to the interrogator's persuasive tactics.

The concept of suggestibility has a long history in psychology: in 1900, Alfred Binet wrote *La Suggestibilité*, and since then books and journal articles have regularly appeared in the psychological literature. In the context of an interrogation, interrogative suggestibility is the extent to which people come to accept as true the messages conveyed by the interrogator during formal questioning. Thus, both publicly and privately, the suggestible suspect agrees with the account offered by the interrogator, even though it differs from their own memory for events.

Some individuals will be more susceptible than others: Gudjonsson and his colleagues have shown that interrogative suggestibility is related to poor recall of the original event (perhaps because of intoxication at that time), acquiescence during questioning, low self-esteem, low intelligence, high emotionality and anxiety, and a high need for social approval.

The picture that emerges from the research is that, through an interaction of situational cues and suspect characteristics, the likelihood of a false confession can be increased.

Psychology in the Courtroom

The psychological study of courtroom processes covers a great deal of ground. This aspect of forensic psychology spans both criminal and civil courts; incorporates diverse areas of law, including juvenile and family law, mental health law and discrimination law; and considers the problem of competency to stand trial. However, perhaps the area in which the study of the courtroom is best seen lies in the study of the jury.

The jury is a natural arena for psychological research: the jury must listen to and evaluate evidence; each juror must formulate an opinion; then the jurors must discuss and reach a collective decision. The cognitive and social processes involved are both intriguing and critical for justice to be done. The problem lies in conducting the research: for obvious reasons, it is not feasible that studies of jurors are conducted during trials, nor in some jurisdictions is it permissible to interview and seek to debrief jurors after a trial. Thus, a great deal of our knowledge of jury and juror behavior comes from, so-called, 'mock jury' studies in which controlled simulations of trials are conducted and 'mock jurors' can be closely observed while key variables are introduced into proceedings.

Extraevidential influences

While evidence presented in the courtroom should inform juror decision-making, there is concern that jurors are subject to extraevidential influences. There are three likely influences: pretrial publicity; witness confidence; and juror sentiments and prejudices.

Pretrial publicity In typical studies, mock jurors are

shown newspaper cuttings containing details about, say, a defendant's criminal record or their retracted confession. When compared to mock jurors who read newspaper stories that did not contain the biasing details, the 'exposed jurors' are more likely to return a guilty verdict after a mock trial. Of course, judges can instruct jurors to ignore pretrial publicity, but this is probably easier to say than to do. In practice, the selection of jurors not exposed to the pretrial influence is the optimum strategy; again, this is probably easier said than done.

Witness confidence Is the confident witness more likely to be believed than the uncertain witness? The weight of evidence from experimental studies suggests that mock jurors are more likely to return a guilty verdict following testimony from confident witnesses. Conversely, jurors place less credibility on the testimony of the less confident witness. Given, as noted previously, that witness confidence is not always a reliable index of accuracy, a belief in a positive relationship between confidence and accuracy is of concern.

Interpersonal perception Are jury decisions and views of the evidence influenced by emotion, sympathy and prejudice? A body of research has pointed to several sources of influence. For example, the perceived attractiveness of the defendant can lead to a favorable outcome, but an attractive defendant with no justification for their offense may be seen in a more severe light. Similarly, defendants of high socioeconomic status are seen as less blameworthy for their offence; while a gender bias suggests that females are less likely to be found guilty for reasons of insanity. Finally, children are seen to be less reliable witnesses in court than adults (which may not be the case).

Some researchers have attempted to estimate the degree of influence these interpersonal factors have in real cases. The evidence actually gathered from real jurors strongly suggests that jurors' decisions are primarily influenced by the evidence, and less so by the perceived characteristics of victims and defendants.

Jurors

Selection There are formal selection criteria, such as age and eligibility to vote, for jury service. The issue here is whether, given these basic formal criteria, all members of the public are equally suited for jury service. The case might be made, for example, that individuals prejudiced against certain groups would not make impartial jurors in given trials. It would be prudent therefore to have the capacity to 'screen out' unsuitable jury candidates.

In practice, screening procedures are typically concerned with questioning prospective jurors with regard to their exposure to pretrial publicity of the case in hand; forming a judgment with respect to their attitudes towards the particular offense being tried; or discovering if they know anyone involved in the case. Other areas of concern might be the attitudes and prejudices of the potential juror, as well as his or her age, sex and education. The selection of jurors is often an emotive issue, taking a great deal of the court's time, and raising the ethical issue of whether the legal process should use scientific jury selection, informed by psychological assessment. The proponents of scientific selection argue for an empirically based method of selecting jurors; opponents hold that selection is unsafe and contrary to the principles on which the jury system stands.

Competence Jurors must assimilate a range of evidence presented during the trial, and understand the judge's presentation of the issues, before they reach a verdict. Several studies have looked at juror comprehension of the evidence in complex cases and their understanding of judges' instructions, typically finding that jurors can struggle to make sense of proceedings. The implication of these findings is both concern if jurors do struggle for comprehension, and the need to clarify information presented in court.

Reaching a verdict The real process of jury deliberation during a trial is, of course, impossible to research. The nearest that can be achieved is either to observe mock jury deliberations or (where allowed) to interview jurors after a trial. The extant research hints at the group dynamics of structure, process and decision-making by 12 people thrown together in a highly charged situation. From appointing a leader, polling views, debating the evidence, recounting the judge's instructions and seeking to apply the law to the evidence, the jurors will be in a state of continual negotiation. There may be disagreements and conflict, and some jurors will become anxious or angry. It may be that issues become personal, even insulting; some individuals may refuse to revise their opinion; some jurors may even refuse to participate, sitting apart from the others; while subgroups may form. Yet, further, the deliberation may go beyond the jury's brief: jurors may blame the defendant for interfering in their everyday lives; and while sentencing is not their concern, jurors may reach a verdict on the basis of what would happen should they say 'guilty'.

The process of jury deliberation is clearly important. Research suggests almost one-third of jurors do change their views; however, the best predictor of the

eventual verdict is the balance of opinion at the onset of the deliberations.

As is often the case in this arena, it is impossible to know how often juries make wrong decisions. It is the case that innocent men and women have been sentenced because of a wrong jury decision; and it is likely that guilty people have been set free. A comparison of the judge's views with the jury verdict may offer an insight into the frequency with which jurors reach a verdict that is, at least, at odds with an expert opinion. American research has compared the convictions and acquittals of judges and juries, finding that in more than 3500 trials, the judge and jury agreed in 78% of cases. When they disagreed, the jury acquitted in 19% of cases when the judge would have convicted, leaving only 3% of cases in which the judge would have acquitted but the jury found the defendant guilty. English research has reported similar findings: most of the time judges and juries agreed, and in instances of disagreement the jury was more lenient.

There are several possible reasons for juror leniency: a defendant with no history of offending may be given a greater benefit of doubt; the jury may feel sympathy towards the defendant; or, in some celebrated cases, the jury disagreed with the law. Of course, to follow the point made above, the jury may reach an incorrect decision because of a failure to comprehend the high standard of proof needed to reach a decision, or because of a lack of understanding of the evidence or the judge's instructions.

The Strength of the Research

As will be clear from the above, a great deal of research findings have accumulated that are relevant to the psychology of evidence. However, a great deal of this research, perhaps particularly in the area of eyewitness memory, has been carried out in laboratory settings. The conduct of laboratory-based research has obvious advantages for the researcher: the laboratory allows the investigator to control what the observer sees; it allows the investigator to select observers, say by age or sex; and it allows the investigator to control how the process of recall and recognition is structured. Indeed, the issue of control is central to laboratory research. By controlling as many variables as possible, the investigator can comment, with a high degree of confidence, on the effects of the particular variables of direct interest in the experimental setting. Thus, for example, in laboratory settings observers generally recall violent incidents less accurately than nonviolent incidents. The key issue that then arises centers on the concept of 'generalizability'; that is, to what degree does the laboratory-based research inform our understanding of real-life cases? A review of the issue has been highly critical of experimental studies, pointing to the limitations of the laboratory (such as the use of staged crimes and filmed sequences, and observer awareness of the focus of the research) as a means by which to understand the complexities of the task facing a real eyewitness. The authors suggest that the limitations of laboratory research are such as to minimize their realism and hence their applicability to real life. Indeed, this point can be followed by questioning the professional wisdom of psychologists acting as expert witnesses and presenting such findings in real cases.

To support the points made above, a field study looked at a real-life case in which 13 of 21 witnesses to a shooting incident were reinterviewed 4–5 months after the event. The performance of the witnesses varied in a number of ways from that which would have been predicted from the laboratory research, particularly with respect to the effects of violence on memory.

Should the findings from laboratory-based studies be disregarded in favor of field studies? The problem is that field studies are open to criticism on the grounds that are the very strengths of laboratory research; that is, the sacrifice of control in the field setting. Thus, in the field study noted above, it is not known to what extent witnesses conferred with each other, how accurate the missing witnesses would be, and whether media coverage had any influence on what witnesses had to say when reinterviewed.

Is there a solution to this apparent impasse? While it is unlikely that the two positions can be fully reconciled, there are two steps that can be made.

First, with regard to the extant research, it is evident that not all the factors studied have an equal strength with respect to their relationship with the accuracy of eyewitness memory. Three categories of research findings have been suggested: (1) reliable and strong factors that show consistent effects on eyewitness memory; (2) reliable and moderate factors that show effects in some studies but not in others; and (3) weak or noninfluential factors that have little or no effect on witness accuracy. In real-life cases the psychologist acting as an expert witness may wish to temper his or her expert testimony in respect of strong rather than weak factors.

Second, looking to the design of future research, it is evident that the optimum approach is to gather data using a variety of research methodologies, including laboratory studies, case studies, field studies and archival studies. Studies that use a range of methodologies and produce consistent findings will considerably reduce the, quite proper, reservations about the applicability of research.

Conclusion

The historical emphasis in much of the psychological research has been to illuminate the fallibility of human performance, so alerting the criminal justice system to the potential for error. More recently, psychologists have turned to a more constructive task in considering ways to aid the gathering of evidence. In particular, psychologists are studying ways of enhancing the quality of witness testimony through more sophisticated interviewing procedures. The 'cognitive interview', which uses memory-retrieval mnemonics to enhance recall, is one aspect of the developing field of investigative interviewing. For the future, forensic psychologists seem set to give increasing attention to the development and refinement of constructive approaches in applying psychology to law.

See also: **Expert Witness**: Qualifications and Testimony. **Legal Aspects of Forensic Science**. **Psychology and Psychiatry**: Overview.

Further Reading

Bull R and Carson D (1995) *Handbook of Psychology in Legal Contexts*. Chichester: Wiley.
Cutler BL and Penrod SD (1995) *Mistaken Identification: The Eyewitness, Psychology, and Law*. Cambridge: Cambridge University Press.
Gheorghiu VA, Netter P, Eysenck HJ and Rosenthal R (1989) *Suggestion and Suggestibility: Theory and Research*. Berlin: Springer.
Gudjonsson GH (1992) *The Psychology of Interrogations, Confessions and Testimony*. Chichester: Wiley.
Hans VP and Vidmar N (1986) *Judging the Jury*. New York: Plenum Press.
Hastie R, Penrod SD and Pennington N (1983) *Inside the Jury*. Cambridge, MA: Harvard University Press.
Hess AK and Weiner IB (eds) (1999) *The Handbook of Forensic Psychology*, 2nd edn. New York: Wiley.
Kagehiro DK and Laufer WS (eds) (1992) *Handbook of Psychology and Law*. New York: Springer.
Kapardis A (1997) *Psychology and Law: A Critical Introduction*. Cambridge: Cambridge University Press.
Kassin SM and Wrightsman LS (eds) (1985) *The Psychology of Evidence and Trial Procedure*. Beverly Hills, CA: Sage.
Konečni VJ and Ebbesen EB (1986) Courtroom testimony by psychologists on eyewitness identification issues. *Law and Human Behavior* **10**:117–126.
Loftus EF (1979) *Eyewitness Testimony*. Cambridge, MA: Harvard University Press.
Milne R and Bull R (1999) *Investigative Interviewing: Psychology and Practice*. Chichester: Wiley.
Ross DF, Read JD and Toglia MP (1994) *Adult Eyewitness Testimony: Current Trends and Developments*. New York: Cambridge University Press.
Sporer SL, Malpass RS and Köehnken G (1996) *Psychological Issues in Eyewitness Identification*. Mahwah, NJ: Erlbaum.

Q

QUALITY ASSURANCE

QA/QC

W J Tilstone, NFSTC, Largo, FL, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0607

Introduction

This article addresses quality in the forensic science laboratory by considering the following questions:

- What are the necessary elements to ensure quality?
- How does the laboratory put these in place?
- How does it provide objective proof that it has done so?

The necessary elements to ensure quality make up the 'quality system'. The laboratory puts the quality system in place through its 'quality manual' and an 'internal audit program.' The proof that it is implementing its quality system is shown by third-party assessment for 'accreditation.' The quality system is the total of policies, practices and procedures which form the basis for good laboratory practice. The system is captured as a quality manual, and accreditation examines the appropriateness of the quality system and compliance with it.

It is also necessary to know what we mean by 'quality'. There are many definitions, but the one used in this article is:

- The quality of a forensic science service is measured by its fitness for the intended purpose.

This definition has the advantage of letting the provider and the user each have a role in achieving quality. The user is defining the purpose and the provider is responsible for ensuring the fitness for that purpose.

The Quality System

An effective quality system is based on three components: people, technical and documentation. We will look at each of these in turn.

People

To produce quality work needs people who know what they are doing, how to do it, why they are doing it, and how well they are doing it. That is, people quality depends on:

- job descriptions;
- recruitment;
- education, training and competency;
- supervision;
- evaluations.

Job description The purpose of the job description is to capture a definition of the job and to communicate that definition. The content of a typical job description will include:

- an understandable, meaningful title;
- responsibilities and accountabilities;
- job outcomes;
- key tasks to achieve outcomes;
- competencies required to perform the tasks successfully;
- education, training and experience standards to confirm the competencies;
- performance goals.

Recruitment The job description provides the framework within which recruitment can take place. Recruitment can be a complex and frustrating activity. There are two main concerns in regard to quality: what to do if there is no ideal candidate; and whether to recruit to promoted posts from within or by hiring 'new blood'.

These questions were at the root of considerable debate in the United States recently when the Quality Assurance Standards for DNA analysis were being developed by the Federal DNA Advisory Board. The standards require that there be a technical leader who has a graduate degree with credits in certain subjects, including molecular biology and population genetics. The objective is to produce good 'DNA problem solvers' who can ensure the standard of operations and development in forensic laboratories, and match

the sometimes very high ability of experts engaged by the defense.

Most technical leaders in post are also the managers of the section, and many do not have a graduate degree. There were also concerns about the ability of people with a PhD in Molecular Biology, but with no or limited forensic experience, to translate their technical knowledge into the particular application. The DNA Advisory Board standards address this through an experience requirement. The problem in regards to existing staff was solved by establishing a waiver process. However, the real answer is to make sure that the purpose of the technical leader is properly and clearly defined in the job description, and that future recruitment is targeted. All candidates can be measured against the elements of the job description and the best one selected.

Education, training and competency These deal with the ability of people to do the tasks assigned to them. Education provides a platform of theoretical and practical knowledge. Training provides continuous development of the theoretical and practical knowledge at a personal level after appointment, so that the individual can perform well as times and job requirements change. Competency is a measure of the overall capability of the person, arising from education, training and on-the-job experience. Competence can be linked to competencies, the knowledge, skills and abilities required to discharge the responsibilities and accountabilities required by the job.

Supervision Good quality requires appropriate supervision. Supervision should have two parts to it: firstly, coaching; and, secondly, responsibility and accountability. The coaching role is the difference between success and failure in delegation and empowerment. Coaching calls to mind theater or sporting analogies. The role of the coach is to make sure that the cast (or team) members know their roles and are able to carry them out. So it is with supervisors in the laboratory. The coaching role works best when coupled with the responsibility/accountability role. The supervisor has to make clear what are the responsibilities of the staff and how they will be held accountable for their performance.

Evaluations Science is about knowledge. Science uses a process of measuring and testing and using the result to further knowledge. It is just the same with people and quality. The performance of people has to be measured and used to further performance.

There are many ways to carry out performance evaluations. Unfortunately, many laboratories have performance review forced upon them through application of an agency- or government-wide system that is divorced from the quality needs of the forensic science operations. Building performance evaluations into the laboratory quality system provides a way to get around the problem. Usually the issues addressed are similar in the laboratory and in the wider agency. The results of the laboratory evaluation can be used to provide the answers to the agency needs.

Technical

The technical overall aspects of a laboratory, and which contribute to quality, are its methods and its instrumentation.

Methods Forensic science operations make substantial and sometimes conflicting demands on methods used. The requirements for method selection are mostly conservative. Methods should be either ones generally accepted in the field, or ones that are supported by information collected in a scientific manner. That is, the methods should either be well established or should be thoroughly validated. The conflict arises when a request is made for testing which is outside the capability of the laboratory at the time. Maintaining quality provides for three responses to such situations: a method can be developed and validated; the request can be passed to another laboratory which is known to have such a method in use; or the request can be turned down.

General acceptability can be shown in several ways. The method may be in widespread use, have been published and extensively cited in the scientific literature, and possibly have been published as a standard method (for example as an American Society for Testing and Materials (ASTM) Standard). One of the problems of 'general acceptability' is that the method may be outdated and there may be better methods available. For example, there is debate about the use of microcrystal tests for identification of controlled substances. There is a view that molecular confirmation techniques such as gas chromatography–mass spectrometry or gas chromatography–infrared spectroscopy (GC-IR) are much more objective. However, many well-established criminalists are able to point to data they have collected that show that reliable and rapid identification can be obtained from careful and knowledgeable application of microcrystal testing.

Validation is a more difficult process in developing a forensic method than in most other analytical applications. The main reason is the nature of the material tested. It is very variable in composition and can be impure.

The most detailed guidelines for method validation

in forensic science are to be found in the quality assurance standards for DNA analysis prepared by the DNA Advisory Board.

They define validations as: a process by which a procedure is evaluated to determine its efficacy and reliability for forensic casework analysis, including:

- Developmental validation, which is the acquisition of test data and determination of conditions and limitations of a new or novel DNA methodology for use on forensic samples.
- Internal validation, which is an accumulation of test data within the laboratory to demonstrate that established methods and procedures perform as expected in the laboratory.

The standards on internal validation have elements that could be applied to any forensic testing area. For example:

> The procedure shall be tested using known and non-probative evidence samples. The laboratory shall monitor and document the reproducibility and precision of the procedure using human DNA control(s).
>
> The laboratory shall establish and document match criteria based on empirical data.
>
> Before the introduction of a procedure into forensic casework, the analyst or examination team shall successfully complete a qualifying test.
>
> Material modifications made to analytical procedures shall be documented and subject to validation testing.
>
> Where methods are not specified, the laboratory shall, wherever possible, select methods that have been published by reputable technical organizations or in relevant scientific texts or journals, or have been appropriately evaluated for a specific or unique application. (Quality Assurance Standards for Forensic DNA Testing Laboratories, as approved by the Director of the FBI, October 1998.)

The normal approaches of analytical chemistry can be used. Accuracy, precision and recovery studies, evaluated by standard statistical techniques, are important characteristics. A range of substrates/matrices should be tested, and aging and environmental insults must be studied.

Instruments The nature of the use to which results are put, together with the nature of the samples tested and the need to be able to select an approach best suited to the overall case conditions, means that instrument calibration and maintenance must be of the highest order. The starting point is to have an inventory of all equipment, the performance of which can affect analytical accuracy. The inventory can then be linked to a calibration and maintenance program. The effect of this is that the laboratory has a program to ensure that the calibration is done and records of the results. The program itself can be staged.

Taking balances as an example. Each should have an entry on the inventory. The entry could show the following:

- Each time used check zero and standard mass.
- Every 6 months check zero, accuracy and linearity of range of standard masses, and off-center load.
- Annual routine preventative maintenance by contractor, with complete recalibration and certification.

The equipment itself must be suitable for the purpose. The laboratory should be able to show its calibration status. Note that 'calibration' here means:

- The set of operations which establish, under specified conditions, the relationship between values indicated by a measuring instrument or measuring system or values represented by a material and the corresponding known values.

The technique commonly referred to as 'calibration', in which a standard curve is prepared for quantitative analysis, is better referred to as 'standardization'.

Documentation

A quality system cannot exist without proper documentation. The three rules are:

- If you do it, write it.
- If you write it, do it.
- If it isn't written, it didn't happen.

These three simple rules give guidance for writing, complying with and recording compliance with the quality system, as contained in a quality manual.

If you do it, write it Many laboratories preparing a quality system for the first time have problems writing their quality manual. The root of the problem is a failure to appreciate that the laboratory must be doing something right or it would not still be in business and seeking to improve its quality. Another common difficulty is that the laboratory tries to create a perfect manual at the first attempt, thus making what should be a simple task much more complex and daunting than it should be. The third common problem arises when the laboratory obtains copies of other laboratories' manuals and seeks to mimic them.

The best way to write the quality manual is to begin by recording what the laboratory actually does. The language should be kept simple and at a high level.

For example, if the laboratory has developed a routine of only analyzing fiber evidence when the supervisor of the trace evidence section has evaluated the case circumstances and makes a judgment call based on experience, then the manual should state that. Trying to capture the exact basis of the judgment is not necessary, especially in the first attempt at compiling a manual. It is especially important to avoid wholesale adoption of another laboratory's manuals, as inevitably there will be practices, policies and procedures in it that do not transfer well.

Starting by capturing what the laboratory actually does also has the advantage that it will result in a more familiar and understandable set of instructions. If the capture is achieved by a process involving staff, then it will have the further benefit of engaging commitment.

If you write it, do it, and if it isn't written, it didn't happen Having captured the policies and procedures current in the laboratory, the next step is to make sure that they are followed. The best way to make sure that the instructions are indeed being followed is through the use of proformas which record all vital steps. They also provide a database for evaluation and improvement.

The advice 'if you write it, do it' does not mean an unthinking and unchanging subservience to the 'rule book'. If experience following something in the manual shows that it can be improved then it should, and the method or procedure or policy should be updated.

The first entirely new policy to be written should, therefore, be the policy on policies. The policy on policies must include the procedures for making and authorizing change.

The Quality Manual and Internal Audit Program

Quality manual

The quality manual documents what is done, by whom, how, with what resources, and where it is done. It covers policy, procedure and practices. In many laboratories, much of the nontechnical content will be set by the parent agency. The quality manual will contain the laboratory's policies and procedures for quality assurance and quality control.

Quality assurance is:

- All activities and functions concerned with the attainment of quality in the laboratory, including: method development and selection, calibration, staff, procurement, reviews and audits, and subcontracting.

Quality control is:

- The operational techniques and activities that sustain the product or service quality to the specified requirements. In the laboratory, it includes: method evaluation, control of standards, materials and reagents, calibration, proficiency tests, and QC samples.

The manual must address the following:

- A quality policy statement including objectives and commitments by management.
- The organization and management structure of the laboratory, its place in any parent organization, and relevant organizational charts.
- The relationships and responsibilities of management, technical operations and support services in implementing the quality system.
- Job descriptions, education, and up-to-date training records of laboratory staff.
- Control and maintenance of documentation of case records and procedure manuals.
- The laboratory's procedures for ensuring that measurements are traceable to appropriate standards, where available.
- The type and extent of examinations conducted by the laboratory.
- Validation and verification of test procedures used.
- Handling evidence items.
- An inventory of equipment, the performance of which affects the quality of the laboratory product.
- Reference measurement standards used.
- Calibration and maintenance of equipment.
- Physical environment control (such as access, temperature, lighting).
- Verification practices for ensuring continuing competence of examiners, including interlaboratory comparisons, proficiency testing programs and internal quality control schemes (e.g. technical peer review).
- Gaining feedback and taking corrective action whenever analytical discrepancies are detected.
- Monitoring court testimony to ensure the reporting of scientific findings in an unbiased and effective manner.
- Laboratory protocol permitting departures from documented policies and procedures.
- Dealing with complaints.
- Disclosure of information and confidentiality.
- Audits, reviews and updates to the quality system.
- Health and safety.
- Accreditations and policy on subcontracting.

The quality manual is completed by the laboratory methods manual. The methods must show the requirements for sampling, equipment, reagents, safety,

controls and standards, possible sources of error, and a step-by-step description of the procedures.

Each policy or procedure must declare its authority. That is, the title, date of issue, authorization, version number or other identification of currency of the policy or procedure. The quality manual is therefore a living document which gives realistic, authoritative, up-to-date and appropriate guidance. Document control is a vital part of achieving these outcomes. Without good document control, staff will not know whether an instruction is currently valid nor what authority is behind it.

Internal audit program

The internal audit program has two goals:

- to demonstrate compliance with the quality system;
- to provide a basis for improvement of the quality system.

The usual form is the vertical audit, in which all activities in a section are audited at the same time. For example, there may be a program in which the toxicology section is audited, then trace evidence, then DNA. A less common alternative is the horizontal audit, in which all instances of a specific activity are audited throughout the laboratory at the same time. For example, evidence control is audited, then training, then case records.

The audit An audit is defined as:

- A planned and documented activity performed to determine by investigation, examination or evaluation of objective evidence, the adequacy of and compliance with established procedures, and the effectiveness of implementation.

The conduct of an audit is the same whether it is a compliance or quality improvement audit, and whether it is a vertical or horizontal audit.

Planning the audit requires attention to:

- Dates when the audit will take place, advising team members and those being audited;
- Team members should be independent of the area being audited and have had training in auditing;
- Scope what is being looked at;
- Documents what documentation is needed (policies, procedures, records).

Conducting the audit requires that the auditors obtain objective information about compliance with the quality system. This is achieved by:

- Questioning staff.
- Observing methods and procedures.
- Examining facilities, documents, records.
- Reviewing the information.

Above all else, it must be understood by the auditors and those audited that:

- An audit is a fact-finding mission and not a fault-finding safari.

Corrective actions If the audit produces objective evidence of nonconformancy, then the laboratory must deal with these through a corrective action program. A nonconformance should always be described in the words of the quality system standard or procedure. If this is not possible, then although the observation may be an opportunity for improvement, it is not a nonconformance.

Nonconformances should be recorded as corrective action reports (CARs). The CAR should describe the nonconformance and initiate a sequence designed to produce an effective rectification, thus:

- Request for rectification by audit team leader.
- Responsibility for rectification assigned by management.
- Root cause investigated by problem-solving team involving front-line staff.
- Corrective action required identified by team and reviewed by management.
- Verification of effectiveness of corrective action assessed by follow-up audit.

Accreditation

Accreditation in the crime laboratory

The objective proof that the laboratory has an adequate quality system and is following it is achieved through accreditation. Accreditation is defined as:

- The formal assessment and recognition by an impartial competent authority that a laboratory is capable of meeting and maintaining defined standards of performance, competence and professionalism.

The strength of accreditation lies in the involvement of the third-party competent authority to assess the performance of the laboratory.

Accreditation programs

There are two main accreditation programs in place in forensic science. The first is that of the American Society of Crime Laboratory Directors Laboratory Accreditation Board, or ASCLD/LAB. ASCLD/LAB was established by the American Society of Crime Laboratory Directors as an independent incorporated body, with the objective of creating an accreditation scheme for crime laboratories, with the following objectives:

To improve the quality of laboratory services provided to the criminal justice system.

To develop and maintain criteria which can be used by a laboratory to assess its level of performance and so strengthen its operation.

To provide an independent, impartial, and objective system by which laboratories can benefit from a total operational review.

To offer to the general public and users of laboratory services a means of identifying those laboratories which have demonstrated that they meet established standards. (ASCLD/LAB, 1999 Accreditation program.)

The Illinois State Police system became the first accredited facility in 1982. In 1990, the Forensic Science Centre in South Australia became the first non-US facility to be accredited. Today there are about 200 accredited laboratories, including those in Singapore, Hong Kong and parts of Australia.

The second approach is to seek accreditation to the standards for good laboratory practice, as set out by the International Standards Organization in ISO Guide 25. Countries adopting this approach amplify the basic ISO requirements with some degree of field-specific requirements. This has been done in the UK, Australia, The Netherlands and Canada. The program in Australia is worthy of note, as the field-specific requirements were arrived at through a co-operative agreement with ASCLD/LAB. Finally, ASCLD/LAB is in the process of revising its program to become fully compliant with ISO Guide 25.

The ASCLD/LAB accreditation program

Some of the principal features of the ASCLD/LAB program will be used to illustrate how accreditation programs affect the quality of operations in a forensic science laboratory.

The program deals with laboratory management and operations, personnel qualifications and the physical plant. Each area is dealt with by describing a set of requirements, presented in the form of a principle, standards, criteria and a discussion. These are defined as:

Principle For each major section within the three divisions of these standards, a basic statement of principle is presented. Principle is defined as: a basic rule, assumption or quality; a fixed or predetermined policy or mode of action.

Standards The standards are statements which describe acceptable levels of performance, excellence, or attainment in that particular activity.

Evaluation criteria The criteria are objective tests to assess whether the laboratory activity meets the standard. This is often a restatement of the standard in the form of a question which can be answered 'yes', 'no', or 'not applicable'. Criteria are each assigned a number in this manual.

Discussion At the end of each of the three divisions, there is discussion which sets forth the rationale used in the adoption of the standards and provides more detailed explanations of some criteria. (ASCLD/LAB, 1999 Accreditation program.)

Essential criteria in the ASCLD/LAB program The criteria are categorized as essential, important or desirable. Essential criteria are those which directly affect and have fundamental impact on the work product of the laboratory or the integrity of the evidence, and only these will be considered further in this article. The original program can be obtained from ASCLD/LAB (ASCLD/LAB, 1999 Accreditation program).

Policies and procedures The first essential criteria encountered are requirements for clearly written and well-understood policies on handling and preserving the integrity of evidence, and on laboratory security. These set the theme for the whole program, in that it requires policies for good laboratory practice as applied to a forensic setting, and then goes on to cover observation of these. Thus criteria rated as essential are next encountered in a set applying to evidence integrity. The policy on evidence integrity should provide the basis for compliance with these operational requirements.

Evidence integrity The first of the group dealing with evidence integrity is that the laboratory has to show that it has a written or secure electronic chain-of-custody record with all necessary data to provide complete tracking of all evidence. Next in the set is a requirement that all evidence is marked for identification. Neither of these is foreign to a forensic laboratory and neither causes much difficulty in interpretation or compliance.

They are followed by criteria requiring that evidence is stored under proper seal and is protected from loss, crosstransfer, contamination and/or deleterious change. Interpretation of these criteria requires careful attention to the wording in the Discussion. There we find a definition of 'proper seal'. A container is 'properly sealed' only if its contents cannot readily escape and only if entering the container results in obvious damage/alteration to the container or its seal. Tape used to seal containers must be initialed (or otherwise identified) to document the person sealing the evidence. The proper seal provision only applies to evidence in storage.

Evidence in the process of being examined is

covered by the criterion requiring it to be protected from loss, crosstransfer, etc. This means that it can be kept in packaging that is not sealed but which must be closed. It also means that it must be kept safely, including consideration of environmental temperature, and security of access. The access issue is addressed by the last of this set of essentials, requiring that there is secure storage space available for overnight and/or long-term storage of evidence.

Quality system The program requires that the laboratory has a quality manual, a quality manager, completes audits of the entire laboratory operations annually, and conducts an annual review of its quality system.

The review of the quality system should address the results of the audit and any other significant factors, such as legislative change, technology developments and issues identified in corrective action activities.

Procedures Procedures used in the laboratory should be generally accepted in the field or should be supported by validation data collected in a scientific manner. For new methods, the validations must be completed before the technique is used on casework. The degree of validation depends on whether the method is being developed anew or is the introduction of an established method not previously used by the laboratory. In the latter case, the object is to show that the method works in the target laboratory, not that it is intrinsically valid – this having been previously established. Therefore, successful completion of samples in an interlaboratory study is acceptable validation.

All the procedures must be documented, and available to staff for consultation. The methods must specify the controls and standards to be used, and the laboratory must be able to show that the quality of the standards and reagents it uses is adequate for the procedure to be used. Finally, in this section, the laboratory must implement a program which checks the reliability of its reagents. This can be as simple as ensuring that test methods require the running of positive and negative controls before case samples.

Instrumentation Instruments must be properly calibrated. Typically this would be a cascading program, with an annual maintenance and calibration against a nationally traceable standard, supported by less stringent calibration checks at regular intervals.

Case records The program has very demanding standards in regard to case records. The reason is to ensure that their integrity can withstand any attack. It is required that notes, worksheets, photographs, spectra, printouts, charts and other data or records used by examiners to support their conclusions are maintained in a case record. The information must be kept as pages, each of which bears the case number or other unique identifier and the identity of the examiner or person responsible for the information. These labeling requirements encompass all documentation in the record, including administrative information. The requirement ensures, among other things, that it is possible to trace the authenticity of any extracts which may have been made from the file.

The records must also be proof against any allegation of alteration. It is thus not permitted to make any alterations in the form of obliterations (such as the use of correcting fluids). Rather, alterations must be made by a single strikeout which leaves the original record readable, and should be initialed by the person making the change.

The laboratory is also required to have a system of review of case records. There are two types of review. Administrative review covers typographical errors and completion of administrative documentation. It is required for all cases and may be performed by the analyst. Technical review covers the correctness of the technical information, including calculations and data transfer, and has to be done by a second person sufficiently knowledgeable of the area to know that the technical steps have been performed correctly and that the data support the conclusions drawn. Technical review does not need to be carried out on all reports but it is suggested that the minimum is 20%, or six per analyst per month, whichever is the lesser.

Testimony monitoring and proficiency testing The laboratory is required to monitor the testimony of examiners at least annually and to provide feedback. The monitoring can be done in several ways, including observation, feedback from attorneys and transcript review. It should cover appearance, poise, performance under cross-examination, ability to present scientific information in an understandable way, and a determination that the testimony is consistent with the examinations and report.

Proficiency testing is dealt with as a set of somewhat complex standards and criteria. Each area of activity must successfully complete at least one external proficiency test per year. The test must be obtained from an approved external supplier and the results passed to the accreditation program for review. Each analyst (other than in DNA) must successfully complete a proficiency test each year, but this can be internally sourced – for example, by re-examination of case samples.

The requirements for DNA analysts are set by external regulation over and above the base require-

ments of the accreditation program. They require each DNA analyst to successfully complete an external test at intervals not exceeding 180 days. If the analyst performs restriction fragment length polymorphism (RFLP) and polymerase chain reaction (PCR) testing then he or she has to complete at least one proficiency test in each.

Corrective action The laboratory is required to have and implement a corrective action program to deal with significant technical problems.

Personnel qualifications Examiners in controlled substances, toxicology, serology, DNA and trace evidence are required to have a baccalaureate degree in an appropriate subject. Examiners in firearms and toolmarks, latent prints and questioned documents do not need a degree. All examiners must demonstrate that they understand the instruments, methods and procedures used.

Competency testing All new examiners (newly recruited or established examiners moving to a new discipline) must successfully complete a competency exam before being allowed to conduct independent casework. The competency test should cover theoretical knowledge (including relevant literature) and demonstrate practical skills, and include a moot court.

Technical support personnel must be able to demonstrate that they meet the requirements of their job descriptions, that their job descriptions and duties agree, and that they have successfully completed a competency test if they have casework responsibilities.

Access control The program requires that all exterior entrance and exit points have adequate security control, and that all internal areas requiring limited or controlled access have a lock system. All keys, magnetic cards and other devices used for access must be accounted for and have controlled distribution. The laboratory must be monitored out-of-hours by alarms or security patrols.

Summary

Forensic science laboratories have a particular responsibility in regard to quality. The results of their work have far-reaching consequences – life or death in some jurisdictions. The nature of the materials that they examine makes quality control much more difficult than in most instances of application of scientific analyses.

This article has illustrated some of the more significant factors that must be considered in quality assurance of a forensic science service.

See also: **Accreditation of Forensic Science Laboratories**, **Administration of Forensic Science**: An International Perspective; Organization of Laboratories. **Education**: An International Perspective.

Further Reading

Information about the ASCLD/LAB program is available from the Executive Secretary, 146 Nicklaus Drive, Garner, NC, 27529. Information is also available from the ASCLD/LAB web site. Information on quality systems and ISO-based programs can be found by visiting the A2LA web site.

Questioned Documents *see* **Document Analysis**: Analytical Methods; Document Dating; Forgery/Counterfeits; Handwriting; Ink Analysis.

S

Scalds *see* **Causes of Death**: Burns and Scalds.

Scene of Death *see* **Causes of Death**: Scene of Death. **Crime-scene Investigation and Examination**: Criminal Analysis; Scene Analysis and Reconstruction.

Self-Inflicted Injury *see* **Clinical Forensic Medicine**: Self-inflicted Injury.

Semen *see* **Clinical Forensic Medicine**: Sexual Assault and Semen Persistence.

Separation Techniques *see* **Analytical Techniques**: Separation Techniques.

SERIAL KILLING

S Hodge, The Centre for Investigative Psychology, Liverpool, UK

Copywright © 2000 Academic Press

doi:10.1006/rwfs.2000.0746

Introduction

Serial murder is a specific type of violence that falls into the crime category of multicide. Multicide distinguishes between the acts of serial murder, mass murder and spree murder. Serial murder is typically defined as a series of three or more separate homicides, with a cooling-off period between each offense, committed at separate locations and involving separate events.

The popular image of serial killers is of evil, disturbed males who roam over wide geographical areas, randomly preying on young women and men. Typically, such definitions are at best oversimplistic, and at worst totally unrealistic. In defining serial murder, discrepancy often occurs as to the number of victims any given offender must kill in order to be termed a serial killer. This important element has been the subject of much debate, and a universally accepted defining number is yet to be found; however, many researchers agree that a serial killer must have a minimum of three to four victims, with the most common number being three victims. Furthermore, many suggest that to include all types of serial killers, the definition must be as broad as possible.

Characteristics of Serial Murderers

The 'typical' serial killer

From an examination of the research into known serial murderers, it is possible to suggest a number of common characteristics that appear to describe a 'typical' serial murderer. Firstly it is well known that the majority of serial murderers are male and that their victims are usually female. Furthermore, serial

killing is characteristically perpetrated by white males on white females, with interracial murder being rare. Offenders are typically aged between 25 and 35, which is older than has been found for the more general crime of homicide, and rarely is there a preexisting relationship between an offender and his victims. All serial murderers are motivated to kill and, typically, that motive is intrinsic to each individual murderer; however, it is accepted that sex is often the most likely motive for many serial killers.

African-Americans

While a large number of known serial killers conform to the profile described in **Table 1**, there is an important group of killers who come from different ethnic groups. In the USA it has been found that 13–16% of serial murderers come from African-American roots. Between 1971 and 1990, there were around 100 cases of serial murder in the USA, 13 of the perpetrators of which were African-Americans. Perhaps the most infamous of such killers is Coral Watts, who has been linked to over 20 sexually sadistic murders between 1974 and 1982 in the states of Michigan and Texas.

Victim choice

It is generally accepted that, in most cases, victims are selected because of their vulnerability; in other words, people who are easier to dominate, such as prostitutes, hitchhikers, the elderly, children and young women. Such people are likely victims not only because of their vulnerability but also because of their accessibility. For example, prostitutes and hitchhikers, by the very nature of their activities, may find themselves willingly getting into the offender's vehicle, and consequently under his control.

In addition to these victim groups, there are also cases where male offenders murder male victims in homosexually motivated murder (Jeffrey Dahmer, Milwaukee, USA; Dennis Nilson, London, UK). Individuals such as homosexual men, prostitutes and runaway children typically lead high-risk lifestyles and as a result are likely to provide more opportunities for offenders. While this opportunity factor is important in determining who becomes a victim, a potential victim may also have to possess particular physical characteristics before becoming an actual victim. For example, Jeffrey Dahmer cruised downtown Milwaukee exclusively for young, black, homosexual men. In other words, not every vulnerable person who crossed his path became a victim.

Table 1 Common characteristics of serial murderers

Characteristic	*Common characteristic*
Offender gender	Male
Offender age	25–35 years
Offender race	White
Victim–offender relationship	Stranger
Victim type	Adult, white, females
Mobility	Local
Motive	Sexual/power/control

Female killers

Another important group to consider is female serial murderers. While female serial murder is a rarity, there are a number of known murderers, who differ considerably from their male counterparts. Of those which fall into this category, many are 'quiet' killers, often working in a caring or nursing environment, such as Beverly Allitt who murdered four children while working as a nurse. She was said to be suffering from Munchausen by proxy syndrome. Others kill as part of a team, usually with a male associate; for example, Myra Hindley and Ian Brady, and Fred and Rose West.

In terms of the differences between male and female killers, female murderers are more likely to use poison to kill their victims. Many, but by no means all, known female serial murderers appear to have been primarily motivated by financial gain. Furthermore, many have been found to have histories of child abuse, extreme poverty and unstable relationships.

Team killers

While serial murderers are typically thought to operate alone, there are a substantial number of cases where murders are committed by two or more individuals together (20–30% of serial murderers are thought to have at least one partner). Some serial killers operate in pairs in order to overpower the victim more easily. Known serial murder teams include Henry Lee Lucus who was helped by his homosexual partner Ottis Toole, and Kenneth Bianchi who killed with Angelo Buono, his cousin. Where the team is made up of a male and a female, the male is often the 'leader' and exerts some sort of psychological control over his partner. Together, Douglas Clark and his girlfriend, Carol Bundy, carried out a number of prostitute murders. A commonality often found between such serial murder teams has been their control of their victims through extreme sexual domination.

Prevalence

Serial murder is not a recent phenomenon: cases have been documented as far back as the fifteenth century (Gilles de Rais, who killed several hundred children in France) but there is clear evidence that the phenomenon has increased considerably from 1970 to the

present day. It is important to note, however, that this recent increase in serial killer rates is actually a direct reflection of the recent increase in the general homicide rate. Nevertheless, serial murder is on the increase and estimates of the number of serial murderers active in the USA at any one time typically range from 35 to 70. Serial murder is not unique to the USA. Although the incidence of serial murder is far higher in the USA, a similar increase over the last 30 years has been documented in many other countries. In the UK, France and Germany it is estimated that 3–5 serial murderers may be active at any one time, while notable cases have also occurred in Russia, South Africa and Australia.

Explanations for Serial Murder

The cause of serial murder is usually tied to one of three explanations: biological, sociological and psychological; however, it is unlikely that any one theory provides a universal explanation of what causes an individual to become a serial murderer. A complex set of factors and influences mold any one individual into a serial murderer.

Biological explanations

Stereotypically serial murderers are seen as crazed maniacs who are biogenetically predisposed to murder. There is a substantial field of research that addresses biological explanations for murder, ranging from the shape of the head and body to chromosome studies. Trauma, brain damage or genetic traits may cause abnormalities in the brain from birth. Occasionally, there will be serial murderers who, when examined, display abnormalities in their genetic make-up. However, as a rule it is unlikely that biological factors can be given as the sole link between humans and violent behavior. In other words, a blow to the head in childhood or defective genes cannot explain all serial murder.

Psychological explanations

The principal explanation given by psychology and psychiatry for the violent personality is that of the psyhopathic personality. The psychopath can be described as possessing specific behavioral traits, including intelligence, a lack of emotion, insincerity and a lack of remorse, and being without the capacity to love.

Some believe that serial murderers are always psychotic. Certainly some serial sexual murderers, for example Joe Kallinger, have been found to be psychotic. Psychotic individuals experience complete breaks from reality, often seeing visions or hearing voices ordering them to kill.

More recently, disassociative disorder, ranging from daydreaming to complete loss of memory, have been used to explain serial murder. There is also evidence for ‘a divided self’, where part of the personality becomes separate to the offender.

In summary, traditional psychology combines biology with psychotherapy to develop its theories. Typical of such explanations are feelings of sexual and personal inadequacy and repressed rage. The principal flaw in the psychiatric approach is that it neatly places people into categories with vague labels, such as ‘psychopath’. Furthermore, the labels themselves are subject to a great deal of disagreement between psychologists and psychiatrists, and there is no explanation of why sexual and serial murder rates vary from one culture to another.

Sociological explanations

More recently, cultural theories of serial murder have evolved, viewing serial murder as a product of learned behavior and cultural violence. In social structure theory, it is proposed that particular groups of people are more prone to criminality because of their social status. For example, poor people are assumed to be more criminal because poverty prevents their attempts to climb the social ladder. As an explanation, social structure theory is unhelpful, as not all serial murderers come from poor backgrounds. Individuals from all social groups may become serial murderers.

In contrast, social process theory assumes that criminal behavior is a result of the process of socialization. In other words, violence is a behavior learned through exposure and normalization from a young age; however, not all children who are brought up in a violent household end up as serial murderers.

In essence, the sociological approach sees power – having it or not having it, obtaining and maintaining it – as the key element in the explanation of serial murder.

Categorizing Serial Murderers

Much of the information and research into criminals is based on taxonomies or classification systems. Although serial murder is believed to represent a relatively small proportion of all homicides, a number of attempts have been made to classify this group of offenders. Consequently, various typologies of both serial murderers and their patterns of offences have emerged; unsurprisingly, many of them conflict with one another. Some are descriptions of causation; some are based on motivation; and others on murder actions and victim–offender interaction.

Motivation

In serial murder, an obvious motive is often missing. In one empirical study of 203 serial murderers, sexual motivation was the most common, although only 20% gave it as their sole reason for killing. Many also listed enjoyment as a motive. It is generally accepted that serial murder is often a reflection of the desire for ultimate power and control over other human beings. The sexual element present in so many serial murders is an instrument used by the killer to obtain that power and control. One interesting classification system distinguishes between four types of murderer based on their apparent motivation for killing.

Visionary killers are compelled to kill because they have heard voices telling them to do so. Such killers suffer from a severe break from reality and are considered psychotic. The motive is always intrinsic and unique to the killer's own personality. The crime scenes are usually chaotic, with a great deal of physical evidence available for investigation.

Mission killers are not psychotic but are driven by a desire to rid the world of a particular group of people whom they see as unworthy of living. From the killing they receive a sense of right because, in their eyes, they have made society a better place. Unlike the visionary killer, the mission killer will plan the murders and will therefore leave little physical evidence at the crime scene.

The hedonistic killer murders for the thrill of it. In this category are both the lust killer and the thrill killer. Both kill for personal and sexual gratification. The difference between them lies in the murder acts they commit. The lust killer kills for sex and the crime will reflect a ritualized sadistic sexual fantasy often involving postmortem acts. The thrill killer also kills for sexual gratification but needs the victims to be alive and aware of the degradation being inflicted upon them.

For the lust killer, the crime scene will show evidence of planning simply because the murder will be a result of fantasy that the killer has played over and over in the mind. There may also be evidence of overkill, torture, strangulation and various postmortem injuries and sexual activity.

For the thrill killer, the crime scene will suggest a high level of planning, as the killer needs to be in a 'safe' location to allow the time needed to humiliate and torture the victim. In fact the crime scene may never be known to investigators, as the thrill killer often disposes of the victims' bodies at a separate location. Unlike the lust and visionary killers, thrill killers will not use overkill. They will be highly controlled individuals and there will be little physical evidence at the scene.

There is a third type of hedonistic serial killer, the comfort killer, who's motivation is material gain. This killer does not kill for sexual gratification, and as a result this type of murder is carried out quickly with no overkill. The crime scene is very controlled and the killer is careful not to leave any physical or forensic evidence.

The final type of murder in this typology is the power/control killer. These killers murder because of a need for complete domination over other human beings. They gain pleasure not from the sexual acts they perform but from knowing that they can do as they wish with the victim. The crime scene will reflect careful planning and control of the situation. Overkill is unlikely and little physical evidence will remain. As with the thrill killer, the victims' bodies will typically be found at a location other than the murder scene.

Serial sexual murder

The most stereotypical of all serial murderers are those who are in some way involved sexually with their victims. It is a widely held belief, by clinicians and law enforcement alike, that the majority of serial murders are sexual in nature. Consequently, in addition to more general typologies of serial murder, typologies focusing specifically on sexual murder have also been developed. For some researchers, the sexual nature of the crime is viewed as a subtype of one or more general taxonomies.

In general, there are believed to be two types of sex murderers: the rape, or displaced anger, murderer and the sadistic, or lust, murderer. The rape or displaced anger murderers rarely report any sexual satisfaction from the murder, nor do they perform postmortem sexual acts. In contrast, for the lust killer, sex plays an integral part in the murder, often when the victim is already dead. The killing fulfills a need for sexual satisfaction and the murders are likely to be ritualistic, reflecting the killer's personal and fatal fantasy.

One of the most widely used typologies of lust murderers is that of the organized/disorganized taxonomy developed by the FBI. The two types are proposed to differ in terms of both crime scene behavior and background/personality traits of the killers. Disorganized lust killers are disorganized in all aspects of life: work, home, clothing and demeanor. They are often loners with feelings of rejection and little success at intimate relationships. In terms of the act of murder, disorganized killers rarely plan attacks and the victim is simply an object for violence. Organized offenders are neat in everything they do and are typically egocentric people who are superficially manipulative and charming. They are less geographically constrained and the crimes are planned and

committed with some expertise. **Table 2** lists the crime scene behaviors and background characteristics that were found to distinguish between the two types.

Mobility

The geographical mobility of serial killers is an area where there is little consensus of opinion and lack of empirical research. Typically, serial murderers are thought to be highly mobile individuals who hunt for victims over a wide geographical area. A number of typologies divide the spatial behavior of serial murderers into categories according to the distances they travel to offend.

Unfortunately, classifications such as those in **Table 3** lack some precision, in that no actual distance ranges are suggested to differentiate between the different groups. However, the typologies do draw attention to the vast variations in the size of area over which serial killers may operate. In terms of linking mobility with crime scene behavior, little research has been carried out. It is suggested, however, that disorganized offenders are unlikely to travel very far from home in order to murder, whereas organized offenders are thought to be more mobile and travel considerable distances to kill.

A number of geographical locations are associated with any one murder in a series (e.g. point of encounter, murder scene, body disposal site) and each and every location used by an offender in the commission of a murder is of psychological and investigative importance. While characteristics of the encounter point will play a role in determining whether the murder can be carried out *in situ*, it can be argued that the stage in the murder process at which an offender chooses to dispose of a body may imply something about that offender, both in terms of mobility and rationality associated with any particular offence. Some researchers suggest that the number of crime locations there are before disposal of the body may reflect the degree of planning a serial murderer has invested in the crime. In other words, the more planned and organized the offence, the more locations involved.

Table 3 Typologies of serial murderer's mobility

Mobility classification	*Description of categories*
Travellers	Crossing state boundaries, covering thousands of miles
Locals	Remain within their home state
Place-specific	Do not leave home to kill
Geographically stable	Live in the same area for some time, kill in the same or nearby area and dispose of bodies in the same or nearby area
Geographically transient	Travel continuously, probably to confuse the police, and dispose of bodies in far-flung places
Hunter	Where an offender leaves homes with the specific purpose of finding a victim
Poacher	Where an offender specifically searches for a victim from a familiar site other than home or who commutes into another area to offend
Troller	Where an offender opportunistically encounters a victim during day-to-day activities
Trapper	Where an offender creates a situation that allows an encounter with a victim in an environment over which the offender has control

Table 2 Crime scene and background characteristics of organized and disorganized offenders

Type of serial murderer	*Crime scene behavior*	*Background characteristis*
Organized	Plan Use restraints Commit sexual acts with live victims Show or display control of victim Use a vehicle	Intelligent Skilled in occupation Likely to think and plan out the crime Likely to be angry and depressed at the time of the murder Likely to have a precipitating stress Likely to have a car in decent condition Likely to follow crime events in the media Likely to change job or leave town
Disorganized	Leave weapon at the scene Position dead body Perform sexual acts on dead body Keep dead body Try to depersonalize the body Do not use a vehicle	Be of low birth order Come from a home with unstable work for the father Have been treated with hostility as a child Be sexually inhibited and sexually ignorant, and to have sexual aversions Have parents with histories of sexual problems Have been frightened and confused at the time of the crime Know who the victim is Live alone Have committed the crime close to home/work

Investigation

The investigation of serial murder poses a particularly difficult challenge for law enforcement. Serial murderers aggravate the difficulties found in all police investigations in two ways: linking cases and identifying the offender. These problems arise from the fact that these murderers are especially difficult to detect. If they were easier to detect, the offender would not have continued to murder time and time again.

The term 'linkage blindness' has been used to describe one of the greatest problems in the investigation of a serial murder. Linkage blindness is the lack of coordination and sharing of investigative information between criminal investigation agencies, which may result in crimes perpetrated by the same individual, but in different states or law enforcement districts, remaining unlinked.

Serial murder is typically also a stranger-to-stranger crime, which makes its investigation all the more difficult. Because of this, forensic investigation and criminal profiling play a vital role in the detection of the offender. There is often a wealth of behavioral information left at a serial murderer's crime scene; this can be used to draw inferences about the type of individual who may have committed such a crime. Such behavioral analysis, combined with a thorough forensic investigation and a consideration of the locations of the offenses, can aid the investigation of serial murder.

See also: **Crime-scene Investigation and Examination**: Recording; Major Incident Scene Management; Criminal Analysis; **Criminal Profiling**. **Psychology and Psychiatry**.

Further Reading

Egger SA (1998) *The Killers Among Us: An Examination of Serial Murder and its Investigation*. Upper Saddle River, NJ: Prentice Hall.

Hickey EW (1991) *Serial Murderers and their Victims*. Pacific Grove, CA: Brooks/Cole.

Holmes RM and De Burger J (1988) *Serial Murder*. Beverly Hills: Sage.

Holmes RM and Holmes ST (1996) *Murder in America*. Beverly Hills: Sage.

Ressler RB, Burgess AW and Douglas JE (1992) *Sexual Homicide: Patterns and Motives*. Lexington, MA: Lexington Books.

Rossmo DK (2000) *Geographic Profiling*. CRC Press.

SEROLOGY

Contents

Overview
Blood Identification
Bloodstain Pattern Analysis

Overview

J Ballantyne, University of Central Florida, Orlando, FL, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0551

Ethos

Serology is defined as the study of the composition and properties of the serum component of blood. Despite this inadequate description of what has colloquially become known as forensic serology, many individuals and laboratories use this rubric to describe a practice that more accurately could be described as forensic biology or biochemistry. 'Forensic serology', then, is the application of immunological and biochemical methods to identify the presence of a body fluid or tissue sample encountered in connection with the investigation of a crime, and the possible further genetic characterization of the sample with a view to determining likely donors thereof. For the purposes of this article the genetic

characterization involves polymorphic cellular antigens and proteins (i.e. those that exhibit variable forms in the population) but does not include DNA genetic markers.

The probative significance of biological evidence transfer

The perpetration of a violent crime often results in a number of different types of biological material being transferred in a unidirectional or bidirectional manner between the victim, the perpetrator, the crime scene or the weapon (**Fig. 1**). A genetic analysis of such biological material by the forensic biologist may yield important legal evidence that may associate, or exclude, a particular individual with the crime in question. Such analysis also may aid in the reconstruction of the sequence of events which occurred before, during or after the commission of the crime. Based upon the circumstances of the case, it is incumbent on the investigators (whether law enforcement officials or forensic scientists) to evaluate the potential 'probativeness' of biological evidence that may have been transferred. By 'probativeness' is meant the degree of meaningfulness of the potential information gleaned by examination of a particular piece of biological evidence as it relates to establishing a relevant fact which may be at issue. Examples of good probative evidence would include the finding of the presence of the perpetrator's semen in the vagina of the rape victim or the presence of the deceased's blood on the perpetrator's clothing. Examples of evidence that would possess low or almost nonexistent probative value would include the finding of blood from a deceased person in immediate proximity to the bloodied body itself, or the finding of semen on the inside of a rape suspect's underpants. The probative value of the evidence greatly increases if there is a demonstrated bidirectional transfer. An example of such transfer would be the transfer of the rape victim's menstrual blood on to the rape perpetrator's underwear concomitant with the assailant's semen being deposited on the victim. In some circumstances the analysis of the nature and distribution of the stain is more important than the genetic characterization of it. A tiny spot of blood consistent with a low-velocity blood spatter pattern on an individual's clothing may belie an attempt to explain the presence of the blood as a result of a Good Samaritan act. Thus it is important to evaluate the circumstances of each crime in order to make a rational judgment as to which evidence to remove for analysis.

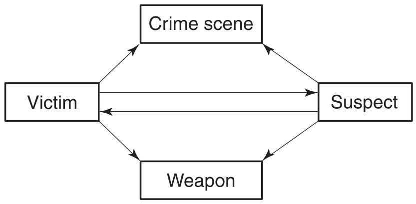


Figure 1 Directionality of biological evidence transfer.

Importance of Communication Between Law Enforcement and Laboratory Personnel

Communication of the circumstances of the case to the appropriate laboratory staff is essential in order that an appropriate analytical scheme can be developed. Failure to do so could result in the inadvertent omission of certain tests that should have been conducted. An example of this would include an altercation that may have been precipitated as a result of one individual spitting at another. If this information is not communicated to the laboratory, it is unlikely that testing for saliva would be performed and an important set of extenuating circumstances might not be corroborated. This is so because testing of clothing items for the presence of saliva is not normally performed in the forensic laboratory unless there is a specific reason to do so.

Importance of Laboratory Search Activities

Evidence submitted to the laboratory must be processed and searched for the presence of probative physiological stains and for other trace materials. Searching takes place under a variety of lighting conditions, both with and without the aid of magnification. This general search phase is critical for locating materials of evidential value, recording information about the nature and location of these materials and collecting and preserving them for later analysis. The general search phase often employs the use of a number of so-called presumptive chemical or biochemical tests for the presence of particular body fluids such as blood or semen. These preliminary tests allow the scientist to screen items efficiently before the use of more specific confirmatory tests.

Identification of Body Fluids and Tissues

Physiological material must first be identified as such before genetic analysis is performed. The most commonly encountered body fluids are blood, semen and saliva, although there may be instances in which the

identification of vaginal secretions, fecal material and urine is necessary.

Blood

Blood consists of hematopoietic lineage cells (red blood cells, white blood cells and platelets) in a proteinaceous fluid known as plasma. Serum is the fluid exuded from blood once it has clotted and thus comprises the plasma minus the proteins (principally fibrinogen and fibrin) responsible for the clotting process.

Presumptive catalytic screening tests The principal function of blood is to transport oxygen to the tissues and remove carbon dioxide therefrom. The hemoglobin molecule (Hb), which constitutes most of the protein content of the erythrocyte, or red blood cell, is principally responsible for this function, and binding of the gas molecules takes place via the nonprotein prosthetic heme group of Hb. The heme group is a nitrogenous planar structure, comprising a protoporphyrin IX ring and conjugated ferrous atom, which happens to possess an associated peroxidase activity:

$$H_2O_2 \rightarrow H_2O + \tfrac{1}{2}O_2.$$

The peroxidase-induced reduction of hydrogen peroxide can be coupled to the oxidation of a number of colorless (reduced) dyes, such as phenolphthalein (Kastle–Meyer reagent), leuco-malachite green (LMG), o-toluidine and tetramethyl benzidine, to form their respective colored moieties. Alternatively luminol (3-amino-phthalhydrazide) can be oxidized to a product that luminesces and this is used to screen large areas for the presence of blood. In this case the area to be searched is sprayed with luminol reagent and the heme-induced luminescence is detected in the dark. The catalytic color tests are extremely sensitive to minute amounts of blood but can produce false-positive results in the presence of a number of substances, including chemical oxidants and catalysts and plant sources containing the enzyme peroxidase. The presence of chemical oxidants and catalysts can normally be excluded by performing a two-step test in which the colorless reduced dye is added first to check for the presence of heme-independent oxidation prior to the addition of the hydrogen peroxide substrate.

Confirmatory tests Subsequent to a positive presumptive test, the presence of blood is confirmed by the immunological identification of Hb or serum proteins, which also serves to identify the blood as being of human (more strictly, primate) origin. Alternatively a positive crystal test for the presence of certain heme derivatives, or the demonstration of the characteristic visible absorption spectrum of hemoglobin, is regarded as conclusive proof of the existence of blood.

Immunological identification of blood and determination of species origin The precipitin reaction occurs when a precipitating antibody combines with its conjugate antigen to produce an insoluble proteinaceous immune complex, which can be detected under appropriate oblique lighting conditions by the naked eye or by the use of general protein stains such as Coomassie Blue or Amido Black. Polyclonal and monoclonal antibodies to serum proteins from humans and a range of domesticated animals are commercially available, as are antibodies to human hemoglobin. The antibody–antigen reaction is either allowed to take place in an inert agar gel by diffusion of the antigen and antibody molecules towards one another (the Ouchterlony reaction or double diffusion in two dimensions) or is facilitated by an electric field (counterimmunoelectrophoresis or crossed-over electrophoresis). An immunochromatographic method, initially developed for clinical use for occult blood, has begun to gain widespread acceptance. In this method a dye-linked mobile monoclonal antibody to human hemoglobin is allowed to react with the sample. If human hemoglobin is present in the sample, it combines with the antibody–dye complex and moves along a membrane until, in a reaction zone, it meets an immobilized polyclonal antibody to human hemoglobin and is concentrated. A positive reaction is indicated by the formation of a dye front. A control zone consists of immobilized anti-IgG that concentrates any unbound mobile monoclonal antibody–dye complex and forms a second dye front.

Confirmatory crystal tests for the presence of blood The two main confirmatory crystal tests are the Teichmann and Takayama tests, named after their developers who initially described the reactions in 1853 and 1912, respectively. These tests rely on the formation of certain heme derivatives. The Teichmann test involves the formation of hematin (heme in which the ferrous ion has been oxidized to the ferric state) halide crystals, whereas the Takayama test induces the formation of salmon pink pyridine hemochromogen crystals. Positive crystal tests, while specific for blood, provide no information regarding its species origin.

Identification of blood by visible spectrophotometry The heme moiety of hemoglobin has a characteristic absorption spectrum which, if present in a sample, is often regarded as conclusive evidence for the presence

of blood. Although different heme derivatives produce different spectra, they all possess the Soret band at 400–425 nm. Oxyhemoglobin produces two absorption bands at 538 nm and 575 nm and a shoulder peak at 610 nm, whereas hemochromogen displays a sharp peak at 550–560 nm. This technique is confounded by the presence in bloodstains of a broad region of absorption at 500–600 nm that tends to obscure the diagnostic heme derivative peaks.

Fetal blood Normal adult hemoglobin is a tetramer consisting primarily of two α and two β polypeptide chains ($\alpha_2\beta_2$). However, during development of the embryo and subsequently the fetus, ε and γ polypeptide chains are respectively expressed instead of the β chain. Fetal blood can thus be distinguished from adult blood by the presence of the fetus-specific γ subunit, which is still detectable up to 6 months after birth. This can be accomplished immunologically with the use of antisera specific to fetal hemoglobin (HbF) or by separation in an electrical field. Specificity problems with many anti-HbF preparations for determining the Hb status in dried and aged stains have prevented its widespread use in forensic laboratories.

Menstrual blood Menstrual blood may be transferred from a female victim of rape or assault to the assailant, and under certain circumstances the identification of it is of some investigative use. Menstrual flow is comprised of endometrial tissue, mucus and blood. Usually not more than 50–60 ml of blood is lost during the uterine cycle and it has the characteristic property of being unable to clot due to extensive degradation of the clotting factor fibrinogen (or its product fibrin). Fibrinogen degradation products are present in relatively high concentration in menstrual blood and can be detected immunologically and normalized against total protein to distinguish menstrual from venous blood. Alternatively, the isoenzymes of lactate dehydrogenase (LDH) can be used to distinguish venous from menstrual blood. Isoenzymes are structurally distinct forms of enzymes that have equivalent catalytic specificities. LDH is a tetrameric protein, the polypeptide chains of which can be of two types, H and M, thus giving rise to five possible isoenzymes H_4, H_3M, H_2M_2, HM_3, M_4 or LDH 1, LDH 2, LDH 3, LDH 4, LDH 5. Venous blood consists primarily of the three isoenzymes LDH 1, LDH 2 and LDH 3, whereas menstrual blood additionally contains elevated levels of LDH 4 and LDH 5. Other tissues possess varying amounts of the LDH isoenzymes and, although the isoforms are readily separated by electrophoresis, difficulties with the presence of body fluid mixtures have limited the efficacy of this technique.

Semen

Semen principally comprises the germ cells (spermatozoa) suspended in a complex mixture of fluids secreted by various accessory glands of the male reproductive tract, including the prostate, seminal vesicles, Cowper's glands and the glands of Littre. Spermatozoa make up approximately 10–25 % of the volume of the semen and normal sperm density ranges from 60 to 100 million per milliliter. The ejaculate volume typically ranges from 1 to 6 ml, with an average of 3 ml, and is dependent upon the time interval since the last ejaculation, the metabolic activity of the glands and the presence or absence of partial ductal obstruction.

Screening tests Screening tests comprise the classical crystal tests for the presence of spermine (Barberio test) and choline (Florence test) and the more commonly used seminal acid phosphatase (SAP).

The Barberio test relies on the formation of spermine phosphate or picrate crystals upon the suspected stain extract's reaction with appropriate anions. The Florence test detects the presence of choline periodide crystals when a semen extract is treated with a solution of iodine in potassium iodide.

SAP is an enzyme present in high concentration in semen; as a nonspecific orthophosphoric monoester phosphohydrolase, it cleaves a variety of organic phosphates, including *p*-nitrophenyl-, α-naphthyl- and thymolphthalein monophosphates. As implied in its name, it is active at acid pH (4.9–5.5). Although SAP is a sensitive test it is not specific for semen owing to its presence in a number of other tissues, including in particular, vaginal fluid.

Confirmatory tests The presence of semen can be confirmed microscopically by the presence of spermatozoa or by the presence of the semen-specific protein p30.

Spermatozoa Spermatozoa have a distinct and characteristic appearance as viewed under the microscope. They are approximately 50–60 μm in length and comprise a flattened ovoid head (dimensions, 4.6 × 2.6 × 1.5 μm) and a 50 μm tail. However, owing to the lability of the tail–head junction, dried stains often possess sperm heads without tails. The head structure, which is principally comprised of a nucleus surrounded by a thin layer of cytoplasm, contains at its anterior end a secretory vesicle known as the acrosome. This appears as a caplike structure and can be differentially stained by standard histochemical stains. The spermatozoa are the principal sources of DNA in semen.

P30 protein P30 protein or prostate specific antigen (PSA) is a protein that is synthesized in the prostate and is an important clinical indicator of malignancy. Its normal range in semen is 300–4200 $\mu g\,ml^{-1}$ with a mean of 1200 $\mu g\,ml^{-1}$. However, it is found in breast, lung and uterine cancers and it may function as an endogenous antiangiogenic protein. Commercial antibodies to PSA are readily available and standard immunochemical techniques can be applied to detect it, including crossed-over electrophoresis, Ouchterlony double diffusion, immunochromatography and enzyme-linked immunoassay (ELISA). The immunochromatographic and ELISA techniques are sensitive to 1 in 10^5–10^6 dilutions of semen (i.e. approximately 1 $ng\,ml^{-1}$) and care must be taken in the interpretation of weak results. For example, it may be possible to get false-positive reactions from postejaculate urine, and urine from adult males, as PSA is present at a mean level of 260 $ng\,ml^{-1}$ therein.

Stability of semen components Dried seminal stains on clothing and bedding can generally exhibit some or all of the semen factors months or even years after deposition. Washing will tend to remove any seminal material, although there have been reports of spermatozoa persistence after machine washing. However, persistence in the postcoital vaginal canal is a different matter and the differential stability of p30, SAP and spermatozoa can be used to assess how long has passed since the last act of sexual intercourse. Semen is lost from the vagina of the living victim due to drainage, dilution with vaginal fluid and phagocytosis of spermatozoa by neutrophilic lymphocytes and mononuclear cells. However, significant levels of p30 tend to be lost within 24 h of deposition in the vaginal vault (as measured by immunodiffusion or crossed-over electrophoresis), SAP is normally lost 48 h postcoitus and spermatozoa do not normally persist after 72 h. In deceased individuals these semen components can last for several days, depending upon the environmental conditions and the rate of atrophy of the body tissues.

Vaginal secretions

Although vaginal secretions are often encountered in postcoital vaginal swabs and stains, there is no definitive test for their presence. The squamous epithelial cells lining the vaginal tract are glycogen-rich and some investigators have used staining with Lugol's iodine to try and distinguish these from other epithelial cells, such as those from the oral cavity. DNA extracts from stains containing a mixture of vaginal epithelial cells and spermatozoa can be differentially enriched for both components.

Saliva

Saliva is a secretion that acts as a digestive aid and contains secretions from the salivary gland. There is currently no definitive test for the positive identification of saliva, although there are a number of substances present in higher concentration in saliva than elsewhere. These include the inorganic anions thiocyanate and nitrite and the enzymes alkaline phosphatase and α-amylase. The presence of significant levels of α-amylase is strongly indicative of the presence of saliva and the detection of α-amylase is the most commonly used test for it.

α-Amylase is produced by two different genetic loci, *AMY1* and *AMY2*. The enzyme hydrolyses the α(1,4) glycosidic bonds of glucose polymers such as glycogen or starch. *AMY1* encodes the salivary form of the enzyme, which is found in saliva, breast milk and perspiration, whereas *AMY2* encodes the pancreatic isoform, which is expressed in semen, vaginal secretions, urine and feces. *AMY1* and *AMY2* can be distinguished by differential inhibition with wheat seed lectin (WSL) and kidney bean extract (KBE), in that WSL and KBE produce greater inhibitory effects on *AMY1* and *AMY2*, respectively. The two most commonly used methods for α-amylase detection are radial diffusion and dyed starch substrates.

Radial diffusion A stain extract from a suspected saliva stain is placed into a well of an agar gel, which also has starch incorporated therein, and allowed to diffuse into the gel. If the extract contains saliva the diffusing α-amylase will hydrolyze the starch and this can be detected by the classical starch–iodine reaction. Starch will give a characteristic purple reaction with iodine, in contrast to a circular clear area where the starch polymer has been hydrolyzed by the α-amylase. A semilogarithmic relationship exists between the diameter of the clear circle and the amount of α-amylase present.

Dyed starch substrates Starch is covalently linked to a dye such as cibachron blue or procion red to form an insoluble complex. Subsequent to α-amylase activity the dye is released into solution and can be measured, for example by spectrophotometry. This forms the basis of the often-used Phadebas test which uses starch–cibachron blue tablets as the substrate.

Urine

Urine contains a variety of inorganic ions, such as sulfate, phosphate and chloride, that can be identified by the formation of their barium, magnesium, ammonium and silver salts, respectively. However, urine also contains a significant number of amines,

including urea and creatinine, and a positive chemical reaction for the presence of amines is regarded as presumptive evidence for the presence of urine.

Urea If the enzyme urease is added to a urine stain it will catalyze the breakdown of any urea present and produce ammonia, which can be detected using a variety of acid base indicators:

$$\text{Urea} + H_2O \overset{\text{urease}}{\rightarrow} CO_2 + 2NH_3$$

Creatinine Urine will give a bright-red coloration in the presence of picric acid and a weak base, the so-called Jaffe reaction for the presence of creatinine.

Amines The general detection of amines is possible by reaction with *p*-dimethylaminocinnamaldehyde (DMAC), which gives a dark-pink/red coloration upon Schiff's base formation.

Fecal material

Fecal material can be identified by a combination of microscopy and testing for the presence of urobilin (which gives feces its characteristic color).

Microscopy Microscopical identification of fecal material relies on the presence of various undigested fibrous food residues, such as meats, fish and vegetables, and enterobacteriaceae such as *Escherichia coli*.

Urobilin In a test known as the Edelman test, urobilinogen (a precursor of urobilin) is oxidized to urobilin by alcoholic mercuric chloride. Subsequent addition of alcoholic zinc chloride produces a green fluorescence, which is due to the formation of a stable zinc–urobilin complex.

Genetic Marker Analysis

Classical genetic markers are inherited biochemical substances that exhibit variation (polymorphism) in the population. Questioned biological stains from the crime scene are typed in various genetic market systems and compared to reference samples obtained from individuals who may be possible donors of the stain. Based upon the results obtained, it is possible either to exclude an individual as being the stain donor or to include that person as belonging to a class of individuals who cannot be excluded as having been the stain donor. Alternative forms of a particular genetic marker are known as alleles, and polymorphic genetic loci are ones for which the most common allele frequency is < 0.95 (or < 0.99). At each of these genetic loci it is possible to assign each individual in the population to one of a small number of possible types, the frequency of which varies according to the alleles present and the subpopulation to which the individual belongs. If there are n alleles then there are a possible $n(n + 1)/2$ possible (geno)-types. Classical genetic marker systems normally possess 2, 3 or 4 alleles, thus giving rise to 3, 6 or 10 subtypes, respectively. Importantly, the more genetic markers tested in a particular crime scene sample, the smaller the proportion of individuals who would possess the constellation of alleles found, and the more probative the evidence. Classical genetic markers can be classified into cellular antigens and extracellular proteins and intracellular isoenzymes.

Cellular antigens

The classical blood groups comprise a diverse set of polymorphic cell surface molecules that are mostly erythrocyte-tethered glycolipids with carbohydrate moieties defining an antigenic specificity that can be detected with an appropriate antiserum. At least 30 different blood group loci exist but only a small subset of these have been successfully employed in forensic serology. These include the ABO, Rhesus and Lewis systems.

ABO blood groups In 1901 Landsteiner discovered that certain combinations of red blood cell suspensions from different people, mixed with blood serum from other people, reproducibly produced a cell clumping or agglutination reaction, whereas other combinations produced no such reaction. Individuals could be classified into four distinct groups, which were named A, B, O and AB, that occur with a frequency of approximately 42%, 8%, 47% and 3%, respectively, in the population. The agglutination reaction takes place because there is recognition of the A or B agglutinogens (antigens) on the cell surface by corresponding isoagglutinins (antibodies). Uniquely, the ABO isoagglutinins are naturally occurring and found in all individuals who are type A (who possess B isoagglutinins), type B (who possess A isoagglutinins) and type O (who possess both A and B isoagglutinins). AB individuals possess neither A nor B agglutinins. A liquid blood sample is typed by firstly separating the red blood cells from the serum and then using commercially obtained monoclonal anti-A and anti-B antisera to test for agglutination of the red blood cells. Confirmatory reverse typing is carried out by testing the sample serum for the presence of isoagglutinins by its reaction with A or B cells. Although the three common ABO alleles (A, B and O) give rise to six different genotypes, AA, AO, BB, BO, AB, OO, only the four types (A, B, O and AB) are distinguishable by liquid typing owing

to genetic dominance effects of the A and B alleles over the O allele. Two common subtypes of A exist, namely A_1 and A_2. Hence there are four common alleles at the ABO locus giving rise to ten different genotypes A_1A_1, A_1A_2, A_2A_2, A_1O, A_2O, BB, BO, A_1B, A_2B, OO.

Secretors and nonsecretors The agglutinogens of red blood cells and the endothelial cells of the cardiovascular system consist of alcohol-soluble glycosphingolipids and, with rare exceptions, all individuals possess these. However, the mucous secretions from the gastrointestinal tract, vaginal secretions and semen of certain individuals contain water-soluble glycoprotein blood group substances with the same antigenic specificity as their red blood cell agglutinogens. These individuals are known as 'secretors' and approximately 80% of the population belong to this group, whereas the remaining 20% of individuals who do not secrete blood group substances into their body fluids are classified as 'nonsecretors'. The secretor/nonsecretor dichotomy has important forensic implications. For example, it is expected that most body fluids recovered from a crime scene (80%) can be ABO typed. In addition, nonsecretor individuals can be excluded as having been the body fluid donor even if the individual's ABO type is the same as that found in the stain, although aberrant secretors exist and each case has to be examined on its own merits.

ABO typing of stains Stains can be typed for the presence of ABO agglutinogens by absorption–elution, absorption–inhibition or mixed agglutination methods, or for isoagglutinins by the Lattes crust method.

1. *Absorption–elution* Blood-stained threads from a questioned stain are affixed by a suitable adhesive to three separate locations on a solid surface. A drop of anti-A is added to the first thread, anti-B to the second thread and anti-H is added to the third thread. If the cognate antigen is present in the bloodstain the antibody (or lectin) will be bound ('absorbed') to the stained thread, whereas if the antigen is absent no such absorption will take place. Upon heating to 56°C, any absorbed antibody or lectin will be 'eluted' and be available for agglutinating red blood cell suspensions of the appropriate type.

2. *Absorption–inhibition* Blood-stained threads are placed in separate test tubes and are able to 'absorb' the cognate antisera, as in the absorption–elution method described above. An aliquot of each antisera supernatant is then removed and tested for its ability to agglutinate red blood cell suspensions of the appropriate type. A reduction in titer of the antisera due to absorption with its cognate antigen results in an inability to agglutinate the red blood cells. Thus 'inhibition' of agglutination of a particular red blood cell type signals the presence of the cognate antigen in the bloodstain.

3. *Mixed agglutination* In this method the antibody is allowed to absorb on to blood-stained threads, as before. Red blood cell suspensions of the cognate antigen are allowed to come into contact with the threads. Any absorbed antibody is detected microscopically by the presence of a layer of red blood cells coating the thread.

4. *Lattes crust* Separate portions of a blood-stained crust are allowed to react with A, B and O red blood cell suspensions. Microscopical observation of agglutination indicates the presence of the cognate isoagglutinin.

Rhesus The clinically important rhesus system consists of six antigens C, c, D, d, E and e, for which appropriate commercial antisera are readily available (except for the d antigen). Rhesus typing of bloodstains can be performed using the absorption–elution methodology but is less sensitive than the ABO system.

Lewis The Lewis system is most commonly used to confirm the secretor status of an individual. Individuals whose red blood cells are Le (a−b+) are secretors, whereas Le (a+b−) individuals are nonsecretors. The rarer Le (a−b−) type is noninformative with respect to secretor status determination.

Extracellular proteins and intracellular enzymes

Genetic variation in the structure of proteins, including enzymes, is mainly due to amino acid substitutions, and at least a third of these are expected to result in charge differences in the protein. In the case of isoenzymes (enzymes which possess the same catalytic specificity but are structurally distinct), such variants are known as allozymes, and they can be electrophoretically separated according to size and/or charge in an electrical field using sieving media such as starch, agarose and polyacrylamide. After electrophoretic separation, isoenzymes are detected using appropriate substrates which produce a color change due to the transfer of electrons from a donor such as NADPH or NADH to a dye such as MTT tetrazolium, thus forming the colored, insoluble MTT–tetrazolium complex. In the case of nonenzyme proteins, the variants are detected postelectrophoretically

using a variety of methods, including immunofixation with human protein-specific antibodies subsequent to capillary transfer of the separated proteins on to an inert membrane (Western blotting). A number of polymorphic isoenzymes and extracellular proteins have been commonly used for forensic purposes. The efficacy of genetic marker systems for individual discrimination can be mathematically ranked by comparing their discriminatory potentials (DP). The DP of a genetic marker system is defined as the probability of discriminating two randomly chosen, biologically unrelated individuals. Mathematically, $DP = 1 - \Sigma p_i^2$, where p_i is the frequency of the ith phenotype. **Table 1** lists some commonly used genetic marker systems and their DP values for a US African American population.

An illustrative example of a genetic marker is the enzyme phosphoglucomutase 1 (PGM1) which is an important 'housekeeping' enzyme involved in the metabolism of glucose. There are four common alleles PGM1*2A(PGM_1^{2+}), PGM1*2B(PGM_1^{2-}), PGM1*1A(PGM_1^{1+}) and PGM1*1B(PGM_1^{1-}), as well as at least 30 rare variants. The enzyme is detected by means of a substrate overlay (zymogram), as shown in **Fig. 2**, where the underlined reagents are those that are added to the 'reaction mix'. The allozymes are separated by isoelectric focusing, which is an electrophoretic technique in which separation is effected in a pH gradient according to differences in the proteins' isoelectric points (the pH at which the net charge on the molecule is zero). The ten common phenotypes are depicted in **Fig. 3**.

Significance of Genetic Marker Typing Data

A typical case involves the genetic marker testing of a body fluid sample from a crime scene and reference samples from the victim(s) and suspect(s). Subsequent to the analysis, conclusions are drawn with respect to possible donor(s) of the identified body fluid. As an illustrative example consider the following genetic marker data obtained from a violent assault, whereby

Table 1 Commonly used protein genetic markers

Genetic marker	*Protein type*	*Discrimination potential (DP)*
ACP1	Isoenzyme	0.67
ADA	Isoenzyme	0.06
AK	Isoenzyme	0.25
CAII	Isoenzyme	0.25
ESD	Isoenzyme	0.27
GLO1	Isoenzyme	0.57
PEPA	Isoenzyme	0.28
GC	Extracellular	0.69
HP	Extracellular	0.62
AHSG	Extracellular	0.54

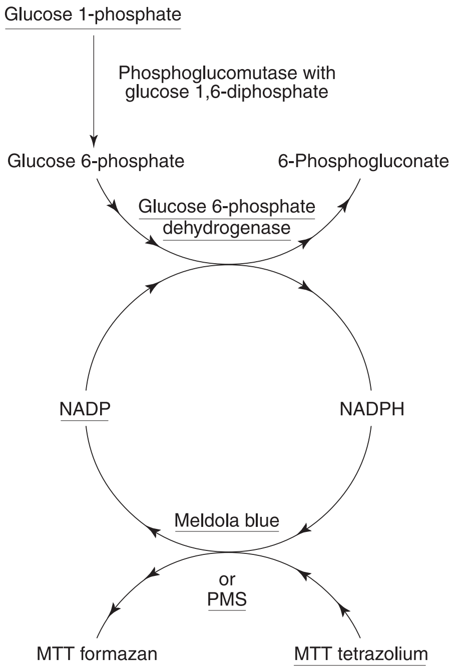


Figure 2 Pathway for phosphoglucomutase reaction. Reagents underlined are those that are added to the reaction mix. PMS = phenazine methosulfate; MTT = 3-[4,5-dimethylthiazol-2-yl] 2,5-diphenyltetrazolium bromide.

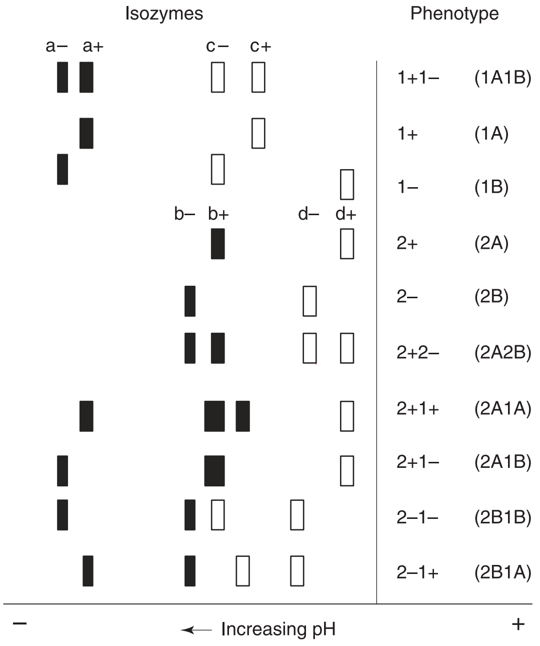


Figure 3 Isolectric focusing patterns of the 10 common PGM1 phenotypes. Black bands indicate increased intensity.

a bloodstain found at the crime scene was typed, as was that of the victim and two suspects:

	ACP1	PGM1	AHSG	ESD	HP	GC
Victim	BA	1+	2−1	1	2	2−1S
Suspect 1	B	2+1+	1	1	1	1S
Suspect 2	B	1+	2−1	1	1	1S
Crime scene blood sample	B	1+	No result	1	No result	1S

Note that it is not uncommon for some genetic markers to yield negative results, as can be seen for the AHSG and HP genetic markers in this case. This can be due to a variety of factors, such as sensitivity of the system and protein stability. Notwithstanding these negative results, what conclusions can be drawn from the above? Firstly, the victim and suspect 1 are excluded as the source of the crime scene bloodstain. Secondly, suspect 2 is included as a possible donor of the bloodstain. What is the significance of this evidence? To evaluate this we ascertain the frequency of each genetic marker type in different population groups from published tables of genetic marker frequencies:

	Caucasian	African American	Hispanic
ACP1 B	0.39	0.57	0.52
PGM1 1+	0.36	0.44	0.29
ESD 1	0.79	0.85	0.77
GC1S	0.31	0.04	0.22
Combined	0.034 or approximately 1 in 29	0.008 or approximately 1 in 120	0.026 or approximately 1 in 39

For example, the ACP1 B type occurs with a frequency of 39%, 57% and 52% in the Caucasian, African American and Hispanic populations, respectively. The probability of obtaining the composite phenotype (i.e. ACP1 B, PGM1 1+, ESD 1, GC 1S) is determined by use of the product rule, by simply multiplying together each system's phenotype frequency. The report would state that suspect 2 could not be excluded as the source of the crime scene bloodstain and that approximately 1 in 29, 1 in 120 and 1 in 39 of the Caucasian, African American or Hispanic population cannot be excluded as possible donors. Both the victim and suspect 1 are excluded as possible sources of the blood.

Sexual Assault Investigations

Conventional genetic marker analysis can aid sexual assault investigations but misinterpretation of the data, whether intentional or not, can occur and a discussion of the issues involved may be useful. A typical sexual assault case involves detection of semen stains on the victim's vaginal samples taken from the victim shortly after the incident. Also as a result of postcoital drainage, seminal stains are often found on the victim's underpants, and such stains can be a good source of seminal material. For the purposes of this exercise let us assume that a woman (victim, V) is raped by an individual (perpetrator, P) but shortly before this she has had consensual intercourse with her husband/boyfriend (B). This scenario must not be uncommon. Before consideration of the scientific data generated from the genetic marker results, a number of mutually exclusive events may have in actuality occurred:

1. Semen is present and *only* comes from P.
2. Semen is present and *only* comes from B.
3. Semen is present and comprises a *mixture* from P and B.
4. No semen is present from either P or B.

From this analysis it is clear that the absence of semen from P does not mean that P did not rape V (due to the possibility of events (2) and (4) above). Other factors may confound the analysis of genetic marker data. For example, semen stains on vaginal swabs (and often on underpants) will be mixed with vaginal epithelia/secretions and therefore testing will identify the genetic factors from V. Although sperm are lost from the vaginal vault (normally not present after 72 h) and seminal stains on underpants are stable (months or years), the vaginal swab best represents semen deposited at the time of rape. The genetic profiles of V, P and B may have alleles in common and masking of one component by another is possible. There is often no scientifically reliable method of determining the number of semen donors using conventional methods. From the foregoing it should be obvious that the interpretation of conventional genetic marker testing data in sexual assault evidence is multifactorial and requires a case-by-case consideration of the meaning of the data generated. The use of highly discriminating DNA typing systems has helped alleviate some, but not all, of these problems. For example, it may be a facile matter to determine the number of semen donors in certain cases with the use of Y chromosome short term repeat (STR) markers.

See also: **Analytical Techniques**: Capillary Electrophoresis in Forensic Biology; Presumptive Chemical Tests; Capillary Electrophoresis in Forensic Science. **Serology**: Blood Identification; Bloodstain Pattern Analysis. **Crime-scene Investigation and Examination**: Recovery of Human Remains. **Deoxyribonucleic Acid**: Future Analytical Techniques; Significance; Statistical Analysis; Mitochondrial Deoxyribonucleic Acid. **Evidence**: Classification; Statistical Interpretation of Evidence/Bayesian Analysis. **Serology**: Overview.

Further Reading

Baechtel S (1988) The identification and individualization of semen stains. In: Saferstein R (ed.) *Forensic Science Handbook*, vol. II, pp. 347–374. NJ: Prentice Hall.

Commissioner of Police of the Metropolis (1978) *Biology Methods Manual: Metropolitan Police Forensic Science Laboratory*. London: Home Office.

Culliford BJ (1971) *The Examination and Typing of Bloodstains in the Crime Laboratory*. Washington, DC: US Government Printing Office.

Divall GB (1985) The application of electrophoretic techniques in the field of criminology. *Electrophoresis* **6**:249–258.

Gaensslen RE (1983) *Sourcebook in Forensic Serology, Immunology, and Biochemistry*. Washington, DC: US Government Printing Office.

Harris H (1980) *The Principles of Human Biochemical Genetics*, 3rd edn. Amsterdam: Elsevier/North Holland.

Harris H and Hopkinson DA (1976) *Handbook of Electrophoresis in Human Genetics*. Amsterdam: North Holland.

Lee H (1982) Identification and grouping of bloodstains. In: Saferstein R (ed.) *Forensic Science Handbook*, vol. **1**, pp. 267–325. Englewood Cliffs, NJ: Prentice Hall.

Race RR and Sanger R (1975) *Blood Groups in Man*, 6th edn. Oxford: Blackwell Scientific.

Sensabaugh G (1982) Biochemical markers of individuality. In: Saferstein R (ed.) *Forensic Science Handbook*, vol. I, pp. 338–403. Englewood Cliffs, NJ: Prentice Hall.

Blood Identification

H C Lee and **E M Pagliaro**, Connecticut Forensic Science Laboratory, Meriden, CT, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0421

Introduction

Bloodstains are often found at various types of crime scene, such as homicide, hit-and-run, assault, child abuse, rape, robbery and burglary. In addition, forensic scientists are often called upon to examine clothing, weapons, vehicles and other items of physical evidence to determine whether a victim's or a suspect's blood has been transferred to those items. In examining blood evidence, the questions which the forensic scientist must answer are: Is it blood? If it is blood, is it human? If it is human, what information towards individualization is possible?

Since the discovery of the ABO system by Landsteiner in 1900, knowledge in human blood identification has expanded tremendously. With the advent of DNA typing technologies in the 1980s an individual bloodstain can now be identified through genetic variation at a molecular level. Further testing procedures allow for greater individualization by DNA typing than was ever possible by classical serological techniques of even just a few years ago. DNA methodologies also allow for the analysis of samples which are highly degraded or present in extremely small quantities. Due to the sensitivity of the polymerase chain reaction (PCR) DNA typing procedures, a minute quantity of DNA recovered from blood, as well as trace amounts of semen, tissue cells, and a variety of body fluids, can yield conclusive typing results. Therefore, it is essential to establish positively the nature of a biological stain and that a stain is, in fact, blood, before conducting further analyses or rendering an opinion concerning the genetic testing result. In addition, certain types of forensic investigations may also require the determination of the species of origin of a bloodstain. In some cases it may be necessary to confirm the presence of human blood in a questioned stain before obtaining a known sample from a suspect or a victim. Thus, it is important to determine if a blood sample is human before proceeding with further genetic testing.

Many techniques have been developed to address these questions. The present state of bloodstain evidence examination is summarized in **Fig. 1**. This article is concerned only with the identification of bloodstains. Discussions of genetic marker typing of bloodstains and the identification of other body fluids can be found elsewhere.

Composition of Human Blood

Blood is a complex fluid tissue composed of a liquid portion, plasma, and cellular components. Plasma is a mixture of dissolved proteins, salts and other chemicals. The blood cells are of three main types: red blood cells (erythrocytes), white blood cells (leukocytes), and platelets (thrombocytes). **Table 1** identifies the major components of blood which can be used for identification purposes.

Erythrocytes are manufactured in the bone marrow. In mammals, the cells lose their nuclei before being released into the circulatory system, and are found as biconcave disks in circulating blood. The average normal red blood cell count is 4.5–5.4 million cells ml^{-1}. Each human red cell is about 7.5 μm in diameter and 2 μm thick. The membranes of human red cells contain a variety of antigens called agglutinogens. Red blood cells function to transport oxygen

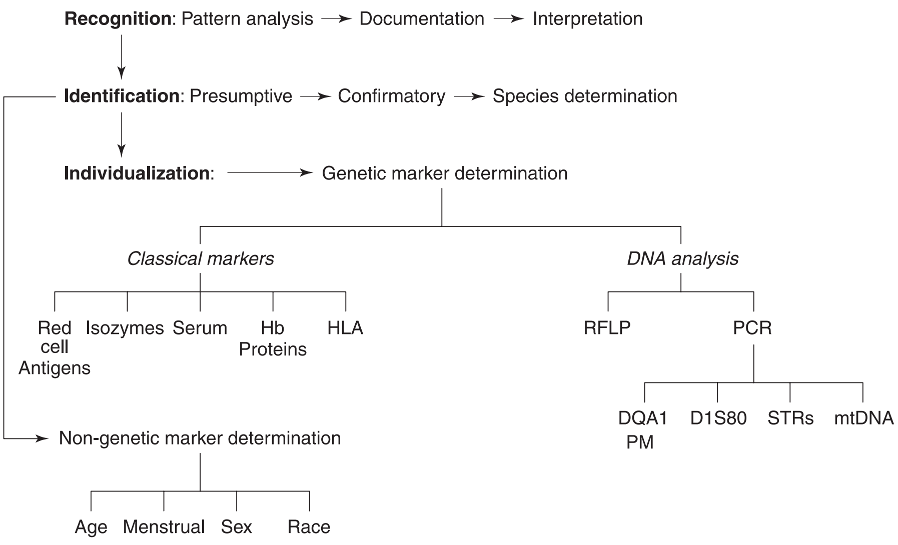


Figure 1 Present approaches to bloodstain analysis. Hb, hemoglobin; HLA, human leukocyte antigen; RFLP, restriction fragment length polymorphism; PCR, polymerase chain reaction; STRs, short tandem repeats; PM, polymarker.

and carbon dioxide and are packed with the complex compound hemoglobin. The pigmented portion of hemoglobin, the heme molecule, gives the red color to the erythrocytes. It is this heme molecule in a stain which is used to identify the presence of blood in a sample of unknown origin.

There are normally 4000–10 000 white blood cells ml^{-1} of blood in the healthy, adult human. White blood cells function as part of the body's defense system and can be of several types. The granulocytes or polymorphonuclear leukocytes are the most numerous. Most granulocytes contain neutrophilic granules (neutrophils); a few contain granules that stain with acid dye (eosinophils), and some have granules that stain with basic dyes (basophils). The other white blood cell types found normally in blood are lymphocytes, cells with large round nuclei and little cytoplasm, and monocytes, cells with abundant cytoplasm and kidney-shaped nuclei. Genetic marker analysis of blood is possible because DNA can be extracted from these nucleated leukocytes.

Platelets are small, granulated bodies, 2–4 μm in diameter. There are about 300 000 platelets ml^{-1} in circulating blood. These cells are involved in the blood clotting mechanism.

Methods for the Identification of Blood

Bloodstain identification is commonly achieved by one of five methods of analysis: (1) microscopic examination; (2) chemical methods; (3) spectrophotometric analysis; (4) immunological methods; and (5) DNA analysis. When using one of these methods, blood is detected by the identification of characteristic components of human blood. These techniques have different levels of sensitivity and specificity, but

Table 1 Components of human blood

Component	*Description*	*Function*
Plasma	Fluid portion of blood: 90% water; 10% solutes, including proteins (primarily albumins, globulins, fibrinogen, and enzymes), glucose, amino acids, lipids, metabolic compounds, respiratory gases, hormones	Transport of soluble nutrients and waste, coagulation
Cellular components		
Erythrocytes (red blood cells)	Biconcave, enucleated cells containing hemoglobin	Transport O_2 and CO_2
Leukocytes (white blood cells)		
Granular	Neutrophils, eosinophils, basophils	Phagocytosis and immune response
Nongranular	Lymphocytes, monocytes	Cellular antibody formation
Platelets	Irregular, fragment-like appearance	Initiate blood clotting

the test which is employed often depends on whether it is being used as a field test or as a laboratory confirmatory test.

Microscopic identification of bloodstains

As liquid blood dries and forms a bloodstain, red blood cells, like other cells, dehydrate. If placed in an environment with a higher solute concentrate, water leaves the cell by osmosis and the cells shrink and change shape. If stains are relatively fresh, it is possible to reconstitute the stain and proceed with microscopical identification of cellular components. A number of techniques have been reported for microscopic examination of erythrocytes and leukocytes in bloodstains. The results obtained with these methods are much affected by the conditions of the bloodstains. Aging, environmental factors, or heating can considerably alter blood cells and make it difficult to produce interpretable and reliable results. In addition to aiding in the identification of a sample as blood, the microscopic appearance of cells found in a stain extract may also reveal other information. For example, sickle-shaped erythrocytes may indicate that the blood sample originated from a person having sickle cell disease.

Reconstituting the blood cells with a solution to restore their original shape can be attempted with the following techniques.

1. Put a fragment of fresh blood crust on the center of a clean slide.
2. Add a drop of a solution of albumin:glycerol: 0.85% saline (20:20:60 by vol.) on the crust and mix gently until the crust is dissolved.
3. Place the slide in a moisture chamber for two hours at room temperature.
4. Prepare a thin film smear of the mixture and air dry.
5. Stain with Wright's stain or other polychromatic stain suitable for blood samples.
6. Examine the slide under the microscope for red blood cells and white blood cells.

Chemical identification of blood

When blood dries to form a bloodstain, the cells are destroyed and their contents released into the surrounding environment. More than 250 proteins, enzymes, and other compounds have been found in the red blood cell, mostly in the soluble portion of the erythrocytes. The predominant erythrocyte protein is hemoglobin (Hb). More than 100 variants of hemoglobin have been described. Identification of blood in stains by means of chemical methods is based on the detection of heme or its derivatives in the stain sample. Such tests can be classified under one of two categories: catalytic tests and crystal tests.

Catalytic tests (screening or presumptive tests) All catalytic blood tests depend on an oxidation reaction in which an oxidant, for example, hydrogen peroxide, oxidizes a colorless material, such as phenolphthalin or tetramethylbenzidine, to a colored one. Alternately, 3-amino-phthalhydrazide (luminol), a colorless material, can be oxidized to a product which luminesces. The general presumptive test reaction is:

$$H_2O_2 + \text{reduced reagent (color 1)} \leftrightarrow H_2O + \text{oxidized reagent (color 2)}.$$

The heme group of hemoglobin exhibits a peroxidase-like activity which may catalyze the breakdown of hydrogen peroxide. The majority of tests which have been devised for the forensic identification of blood are based on the peroxide-mediated oxidation of leukomalachite green, phenolphthalin, *o*-tolidine, luminol, tetramethylbenzidine, fluorescein, and other less commonly used compounds. At one time, benzidine and its derivatives were widely used as the color reagent in screening tests for blood. However, due to the carcinogenic nature of these compounds and the health risks involved in their use, laboratories no longer use these types of chemical reagents. The tests most commonly employed in modern crime scene procedures are phenolphthalin, leukomalachite green, luminol and tetramethylbenzidine. Reaction schemes for some of these common chemical reagents are shown in **Fig. 2**. All of these chemicals are highly sensitive to minute traces of hemoglobin and its derivatives, but all suffer from the occurrence of false positive reactions with some of the following materials: catalases, peroxidases, cytochromes, strong oxidizing agents and metallic salts.

Testing procedures Prior to testing, the nature, color and appearance of the stain should be noted. These are important data which will assist the scientist in interpreting any positive reactions noted with the test reagents. All efforts should be made to limit alteration of the stain or pattern while performing the screening test for blood. The screening test for blood should be performed by scraping a small sample from the stain or removing a small portion of stained material.

1. Color reagent is added to the stain material.
2. If no color develops within 30 s, a drop of 3% hydrogen peroxide is added. A resulting color change indicates that blood may be present.
3. Alternatively, the stain area may be lightly rubbed with a clean cotton swab or filter paper moistened

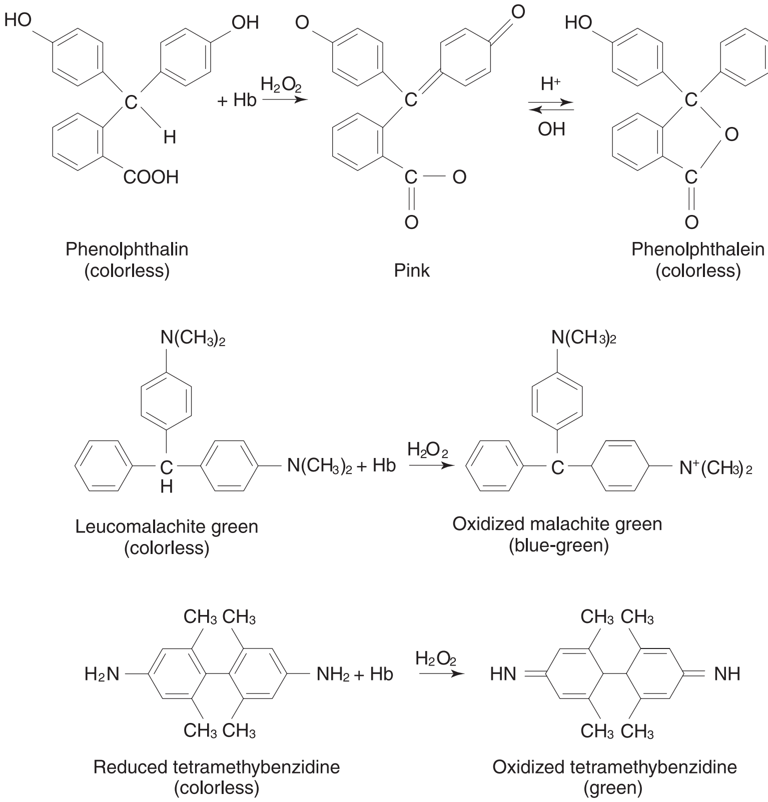


Figure 2 Reaction schemes for some presumptive test reagents.

with distilled water. Reagent and hydrogen peroxide are then added to the swab sample. This method is preferred for crime scene work and when determining which stains warrant additional testing on items of physical evidence.

4. Small samples of suspected blood mixed with other materials, such as soil, may be dissolved and the resulting supernatant tested accordingly.

Interpretation of results Color catalytic tests are very sensitive, but not specific. A positive color test alone should not be interpreted as positive evidence of blood. However, a negative result is proof of the absence of detectable quantities of heme or its derivatives.

When a positive result is obtained, it is necessary to consider carefully whether the test could have resulted from something other than blood. The specificity of various catalytic reagents has been studied extensively. A false positive reaction may be defined as any positive reaction given by any substance other than bloodstains. These substances may be conveniently divided into three groups.

1. *Chemical oxidants and catalysts*: copper and nickel salts are the chemicals which most frequently show false positive reactions. Others are rust, potassium permanganate, potassium dichromate, some bleaches and hypochlorites, iodine, and lead oxides. Among the common test reagents, phenolphthalin gives positive results with oxidizing compounds such as copper, potassium ferricyanide, nickel and cobalt nitrates, and some sulfocyanates; luminol reacts with some compounds of copper, cobalt and iron, potassium permanganate and hydrated sodium hypochlorite.
2. *Plant sources*: vegetable peroxidases are the most important class of substances which show false positive reactions with chemical color tests. The following plant tissues may react with the *o*-tolidine or phenolphthalin reagents and be mistaken for blood: apple, apricot, bean, blackberry, Jerusalem artichoke, horseradish, potato, turnip, cabbage, onion and dandelion root. Plant material, such as horseradish, garlic, cabbage, tomato and cucumber, reacts positively with tetramethylbenzidine.
3. *Animal origin*. the following substances of animal origin may give positive reactions with screening reagents: pus, bone marrow, brain tissues, spinal fluid, intestine, lung, saliva and mucus. These reactions may be attributed in many cases to minute quantities of blood in the tissue or body fluid samples being tested.

Many false positive reactions caused by material other than blood can be eliminated by the following methods.

1. *Chemical oxidants and catalysts*. the behavior of chemical oxidants is quite different from that of blood. Chemical oxidants will give a color change before the addition of the hydrogen peroxide. Therefore, a false positive reaction can be distinguished by use of a two-step test procedure in which the color reagent is added to the sample first. If no color change occurs within 30 s, the peroxide is added. A color change after peroxide addition indicates the presence of blood or other peroxidases.
2. *Plant peroxidases*: heme is stable at high temperatures, whereas plant peroxidases are rapidly deactivated. Therefore, heating the sample stain or extract to 100°C for 5 min will differentiate the plant peroxidases from blood sources. Also, it has been found that peroxidases can be separated from hemoglobin by electrophoretic methods; however, in most instances in a laboratory or field setting this procedure is impractical. Use of some other method of blood confirmation, such as utilization of antihuman hemoglobin is most often employed.
3. *Other substances of animal origin*: microscopic examination of the specimen will distinguish tissue, pus, and other substances of animal origin from blood.

Crystal tests There are several crystal tests that are considered by most forensic scientists as a confirmatory test of bloodstains. Crystal tests are based on the formation of heme derivative crystals such as hematin, hemin and hemochromogen. In 1853, Teichmann reported that by gently heating blood with glacial acetic acid in the presence of salts, crystals were formed. The positive result is due to a combination of a halogen with ferriprotoporphyrin. The crystals are rhombic or prismatic in shape, dark brown in color, and about 10 μm in size. The age of the bloodstain does not affect the formation of hematin crystals. Twelve-year-old stains have given a positive Teichmann. Bloodstains over 20 years old were also reported positive with this test. Takayama suggested a method of producing a heme complex called hemochromagen in 1912.

The sensitivity and specificity of the Takayama and the Teichman tests are similar: the test is positive with as little as 0.001 ml of blood or 0.1 mg of hemoglobin. While Teichman crystals may not form with stains from materials such as leather, older stains have been reported to give false negative results when tested with Takayama reagent. Therefore, the failure to obtain a positive crystal test does not necessarily indicate the absence of blood.

Spectrophotometric methods of blood identification

Spectrophotometric procedures are seldom used at the present time in forensic analyses. This type of examination is based on the identification of hemoglobin and its derivatives through their specific absorption spectra. Absorption spectroscopy of hemoglobin was first described by Hoppe in 1862 as a means for blood identification. During the early days of the development of procedures, this method was considered one of the most conclusive tests for the identification of bloodstains. The determination of near ultraviolet and visible absorption spectra allows sufficient reliability and sensitivity for the identification of hemoglobin and derivatives such as methemoglobin, oxyhemoglobin, carboxyhemoglobin, and sulfhemoglobin.

In the near ultraviolet and visible regions of the spectrum, a complex system of absorption bands is present due to the heme portion of the hemoglobin molecule. The visible region of the spectrum of the heme derivatives differs substantially from derivative to derivative, but all have in common a strong absorption band at 400–425 nm (the Soret band). Porphyrin compounds and their derivatives from other animal or vegetable sources may share spectral characteristics with hemoglobin, hematin, or hemochromogen. Therefore, the identity of bloodstains should never be inferred solely from a single absorption spectrum.

Electrophoretic methods

Two electrophoretic approaches have been recommended for identifying bloodstains: (1) separation and identification of hemoglobin by electrophoresis and (2) separation and identification of serum proteins by immunoelectrophoresis.

Hemoglobins are conjugated proteins. After selection of an appropriate buffer pH, the charged hemoglobin molecules are moved by electrophoresis through a support medium toward the electrode with the opposite charge. Most of the substances which give false positives with chemical tests for blood are either uncharged or have a different charge from hemoglobin and are thus eliminated by this method. After electrophoretic separation, the hemoglobin fractions are visualized by staining with leukomalachite green solution or any other catalytic color test reagent. Banding patterns may then be compared with known standards.

Immunoelectrophoresis involves the combination

of the techniques of immunodiffusion and electrophoresis for the analysis of biological fluids. In this procedure, bloodstain extract is placed in wells in agar on a glass slide and then subjected to electrophoresis by application of an electric current. A bloodstain extract contains hemoglobin as well as serum proteins. Under these conditions, the individual proteins move at different rates. After electrophoresis, antihuman serum is placed in a trough running the length of the slide and parallel to the path of migration. The separated proteins and antiserum diffuse toward one another, permitting the corresponding human serum proteins to undergo an antigen–antibody reaction with the antibodies and forming preciption lines at the points where these complexes form. The hemoglobin will remain near the point of origin and give a pinkish ring around the sample well. These white precipitin lines and the pinkish hemoglobin ring are a positive indication of blood. There are no other substances besides blood that will give this pattern combination. Another advantage of this method is that the species of origin of the bloodstain can be determined at the same time.

Immunological (antihemoglobin) tests

Anti-hemoglobin precipitin sera have been used for the identification of human bloodstains. The highly specific reaction obtained between human bloodstains and the antihuman hemoglobin serum allows a stain to be identified in a single operation as blood of human origin. This test can be carried out through either one-dimensional or two-dimensional diffusion techniques. A positive result with this test is not only absolute identification of a stain as blood, but also shows the stain is from a human source.

Determination of Species of Origin

After a stain has been identified as blood, it is necessary for the forensic scientist to determine whether that blood is of human origin. If it is not human, it may then be necessary to determine to what species the blood does belong. Most methods in common use for determining the species of origin are immunological in nature. If an animal is injected with a protein molecule from another species, it will sometimes recognize this protein as a foreign substance (antigen) and will produce an antiserum (antibody) which will react with such protein both *in vivo* and *in vitro*. The immunological precipitin test for medicolegal species determination in bloodstains was first used in 1901.

The *in vitro* antibody–antigen reaction is detected by the formation of an antigen–antibody (Ag–Ab) complex. The reaction requires the presence of three elements: antiserum, bloodstain extract (antigen) and buffer. The temperature, pH, incubation time, and ionic strength at which a precipitin reaction is performed have a direct influence on the precipitin band formation. For example, the most favorable temperature is usually between 25°C and 37°C and the optimal pH is between 7 and 8. However, the exact conditions which are optimal for a system must be determined for each new antigen–antibody system under investigation.

The specificity of the antiserum plays the most important role in species determination. Traces of contaminating antibodies in commercially prepared antisera could cause serious error. Therefore, the precise specificity of the antiserum in use must be known. Antihuman sera can be produced by injecting human serum or hemoglobin into various animals. The most commonly used antisera created by this method are produced by rabbits, goats or sheep. These antisera produce a stable precipitate. Monoclonal antibodies are also commercially available for species testing. To ensure the specificity of the antiserum, it is imperative that laboratory scientists select by direct testing for crossreactivity and determine the strength of the antiserum by a titration method. During species determination, the same batch of tested antiserum must always be used. Only by such strict controls can the forensic scientist maintain the degree of certainty and reproducibility required for a reliable species determination.

Figure 3 depicts several methods for the determination of the Ag–Ab complex in species tests.

The following are the most commonly used methods for species determination in forensic laboratories.

Precipitin methods

Ouchterlony method: double diffusion in two directions This diffusion method was first described by Ouchterlony in 1949. It involves the use of agar gel plates with wells for both antibodies and antigens. The two reactants diffuse into the gel where the soluble antigens and antibodies form an insoluble complex – a precipitate. The Ouchterlony method allows both qualitative and semiquantitative evaluation of the reactants. Precipitin band formation gives the scientist considerable information regarding the identity, partial identity or nonidentity of the antigen and antibody reaction. It also yields information on the diffusion coefficients and concentrations of the reactants.

Crossed-over electrophoresis The crossed-over electrophoretic technique can be used for both quantitative and qualitative determination of a blood sample.

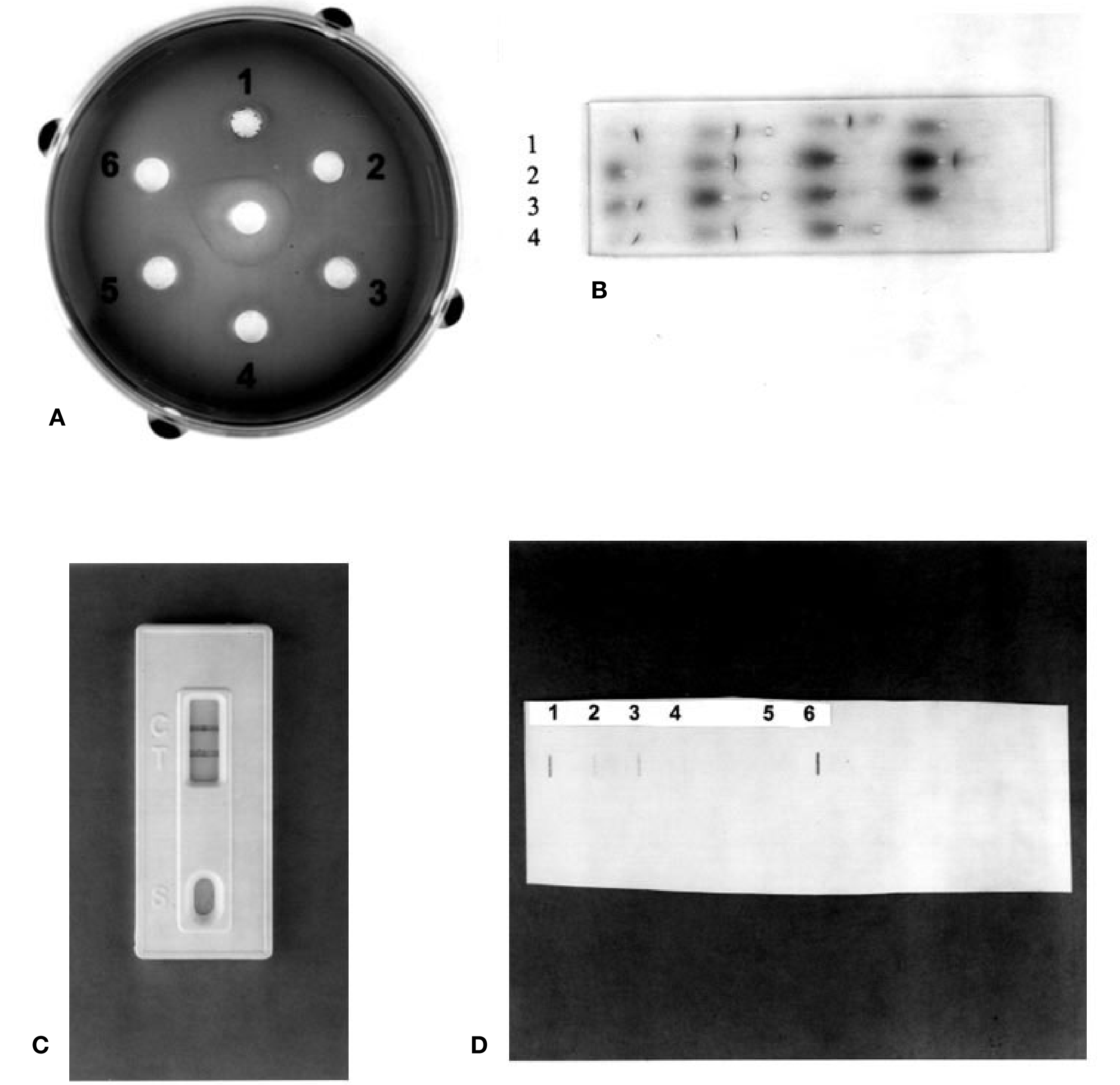


Figure 3 Common methods for species determination. (**A**) Ouchterlony double diffusion: samples 1 through 5 show a positive reaction with antiserum (center well). (**B**) Crossover electrophoresis: samples 1, 3 and 4 are positive; sample 2 is a negative control. (**C**) Rapid immunoassay: T, test sample with a positive result; C, positive control. (**D**) Human DNA quantitation: slot blot showing various positive DNA concentration standards (1–4), negative control (5) and positive test sample (6).

Under the influence of an electric field, the antigen and the antibody migrate toward each other and a precipitate is formed at the point of their interaction. Small wells about 1.5 mm in diameter are punched in an agar gel. The stain extract is placed in the cathodic well of a neighboring pair, and the antiserum in the opposite well. The antiserum travels towards the cathode (negative electrode), while the stain extract migrates anodically (toward the positive electrode). A precipitin band will form at the site of the interaction of the antibodies and the antigens.

Rapid immunoassay

Immunoassay test strips for human blood have become commercially available in recent years. These tests offer high sensitivity and specificity as confirmatory tests for the presence of human blood in stain extracts. Such procedures involve the reaction of antigens in the extract with monoclonal antibodies within the test strip. The combined antigen–antibody complex moves up the strip to an area where it is immobilized. The complex reacts with dye particles in

the area where it has been halted to create visible reactions. Some assays combine built-in positive controls on the test strips. These tests have a reported sensitivity of 0.05 μg hemoglobin ml^{-1} and, thus, are suitable for use with highly dilute stain extracts or older stain materials.

Rapid immunoassay procedures are highly specific, easily performed, applicable to various types of samples, and produce results in a timely manner, usually within minutes of application to the test strips. This last feature allows for confirmation of blood in supernatants of samples which will be subject to DNA analysis prior to testing, without consumption of stain portions necessary for obtaining DNA. Such rapid assays also prevent extensive delay in genetic marker testing. Rapid confirmation of human blood may also be conducted at crime scenes by trained personnel. The disadvantage of this method is that reported 'high dosage' effects may yield false negative results. However, this effect is readily avoided by appropriate dilution of the sample prior to application of the stain extract to the test strips.

Human DNA quantitation

A sample can be determined to be blood of human origin by reaction with a probe specific for human DNA. Probes complementary to primate specific DNA sequences, such as those found at the locus D17Z1, are readily available and used primarily to determine the amount of human DNA extracted from a sample prior to DNA typing. DNA extracted from a sample is spotted on a membrane along with known concentrations of human DNA. After reaction with the human-specific probe, results obtained from the unknown sample are compared to the signal intensity of the known standards; the amount of human DNA in the sample can be estimated in this manner. The sensitivity of the human DNA quantitation test is commonly in the 0.15–0.20 ng range when a color reagent detection method is employed. One disadvantage of this technique is that human DNA from any tissue or cells will produce a positive reaction. Thus, it is necessary to determine that the DNA obtained is from blood by using one of the heme identification techniques discussed previously.

See also: **Analytical Techniques**: Microscopy; Spectroscopy: Basic Principles; Presumptive Chemical Tests. **Serology**: Bloodstain Pattern Analysis. **Analytical Techniques**: Spectroscopic Techniques.

Further Reading

Baxter SJ and Rees B (1974) The use of anti-human haemoglobin in forensic serology. *Medical Science and the Law* **14**:159–162.

Culliford BJ (1964) *The Examination and Typing of Bloodstains in the Crime Laboratory*. Washington, DC: US Department of Justice.

Gaensslen RE (1984) *Sourcebook in Forensic Serology, Immunology, and Biochemistry*. Washington, DC: US Department of Justice.

Gaensslen RE and Camp FR (1981) Forensic serology: analysis of bloodstains and body fluid stains. In: Wecht CH (ed.) *Forensic Science*, vol. 2. New York: Matthew Bender.

Hochmeister MN, Budowle B, Sparks R *et al.* (1998) Validation studies of an immunochromatographic 1-step test for the forensic identification of human blood. *Journal of Forensic Science* **44**:597–602.

Inman K and Rudin N (1997) *An Introduction to Forensic DNA Analysis*. New York: CRC Press.

Lee HC (1982) Bloodstain identification and grouping. In: Saferstein R (ed.) *Forensic Science Handbook*, vol. I. Englewood Cliffs, NJ: Prentice-Hall.

Lee HC and Gaensslen RE (eds.) (1985) *Advances in Forensic Science*, vol. I. Foster City, CH: Biomedical Publications.

Miller LB (1969) Hemochromogen crystal formation with minute amounts of blood. *Journal of the Forensic Science Society* **9**:84–86.

Shutler GG (1980) A Study of the interrelationship between fingerprint developing techniques and bloodstain identification and typing methods. *Canadian Society of Forensic Science Journal* **13**:1–8.

Walsh PS, Valero J and Reynolds R (1992) A rapid chemiluminescent method for quantitation of human DNA. *Nucleic Acids Research* **20**:5061–5065.

Waye JS and Willard HF (1986) Structure, organization, and sequence of alpha satellite DNA from human chromosome 17: evidence for evolution by unequal crossing-over and an ancestral pentamer repeat shared with the human X chromosome. *Molecular and Cellular Biology* **6**:3156–3165.

Bloodstain Pattern Analysis

T L Wolson, Miami-Dade Police Department, Miami, FL, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0422

Introduction

When violent crimes are committed, it is not unusual for the participants to be injured. If these injuries are accompanied by blood flow, distinctive bloodstain patterns may result that can be used to provide

investigative information about the activities which occurred during the commission of the crime. These distinctive bloodstain patterns occur because of the physical properties of the blood and how it reacts when acted upon by physical forces. The analysis of these patterns can provide the investigator with information about the direction of travel of the blood, the level of force used to put the blood in flight, the location of the blood source which was acted upon to create the pattern, movements during bloodshed, movements after bloodshed, and activities during bloodshed. Bloodstain pattern analysis studies how different forces and activities influence the creation and appearance of the bloodstain patterns, so that they can be interpreted as part of the crime scene investigation.

Bloodstain Characteristics

Many factors affect the size and shape of the blood drops. When blood passively drips off of a surface and falls on to a smooth hard horizontal surface, the resulting bloodstain will be round. As the blood falls through the air, it takes the shape of an oscillating sphere. The blood droplets will not break up in the air as they fall due to gravitational force alone. Additional forces, however, can break the drops apart. The diameter of the bloodstain is dependent upon the distance the drop falls to the horizontal surface as well as the volume of the blood drop. As the distance of the fall increases, the diameter of the bloodstain will increase until it reaches a maximum diameter. The maximum diameter for the bloodstain occurs after the drop has fallen approximately 1.8 m (6 ft). The volume of the drop of blood also affects the diameter of the bloodstain. As volume increases, so will the diameter of the resulting bloodstain. Early research identified the average volume of a drop of blood as 0.05 ml. Subsequent research determined that the average volume of a drop of blood varies depending on the surface characteristics of the object that the blood drips from. Blood drops fall off of a surface because its volume increases to the level where the pull of the Earth's gravity overcomes the viscosity of the blood and allows the surface tension to break. The shape and finish of the surface the blood falls from affects the blood volume that is needed for the surface tension of the blood to break and allow it to fall.

In addition to droplet volume, distance of fall and the blood source surface characteristics, other factors affect the size, shape and appearance of the bloodstains. Droplet size is a factor of how much force was imparted on the blood source to put it in flight. The viscosity, specific gravity and surface tension of blood make it resistant to being broken up into drops. When an external force is imparted on static blood, it will cause some of the blood to react and be put in flight. The distance the drops of blood fly through the air is dependent upon how much force was used to create the drops, the size of the blood drops, and air currents. In low force events (sometimes referred to as low velocity), the number of blood drops put in flight is low and the size of the droplets tend to be large. The majority of the droplets will be larger than 3 mm in diameter, with very few having a diameter less than 3 mm. The distance these droplets travel from the blood source tends to be short. As the level of force or velocity increases, the number of blood drops put in flight and the distance they travel away from the blood source increases. The diameter of the drops decreases as force increases, with many of the blood drops being 1 mm in diameter. There will still be larger drops but these drops tend to travel a greater distance away from the blood source than the smaller 1 mm and 2 mm diameter drops. Small drops of blood travel less distance, as they lack the physical weight to resist the air currents and friction, and the energy which puts them in flight dissipates rapidly. The majority of the small blood drops will not travel more than 0.9 m (3 ft) from the blood source. When the level of force becomes extremely high, such as in gunshot spatter, many thousands of blood drops, which are 1 mm and less in diameter, will be put in flight. As distance from the blood source increases, the size of the droplets increases, but these droplets are small (2 mm and 3 mm) in comparison to the large drops created by a low force bloodspatter event. Many of the larger drops, which are created in these high force events, are the result of smaller blood drops colliding while in flight and combining into larger volume drops of blood.

Bloodstain shape is determined by the angle between the flight path of the droplet and the surface that it impacts (**Fig. 1**). When a drop of blood strikes a horizontal surface from the perpendicular (at an angle of 90°, **Fig. 2**), the resulting bloodstain will be round. As the angle between the blood drops' flight path and the target surface decreases, the length of the resulting bloodstain increases and the width decreases; in other words, the bloodstain becomes longer and narrower as the size of the angle decreases (**Fig. 3**). A blood drop that has impacted a target surface from an angle of 12° will be long and narrow (**Fig. 4**). Also present on these long, narrow stains is a 'tail' (Fig. 4). As the drop of blood impacts the target surface, the main body of the droplet will stick to the surface. A small portion of the blood drop will tear off from the top and continue in a forward direction. This secondary spatter has the appearance of a tail.

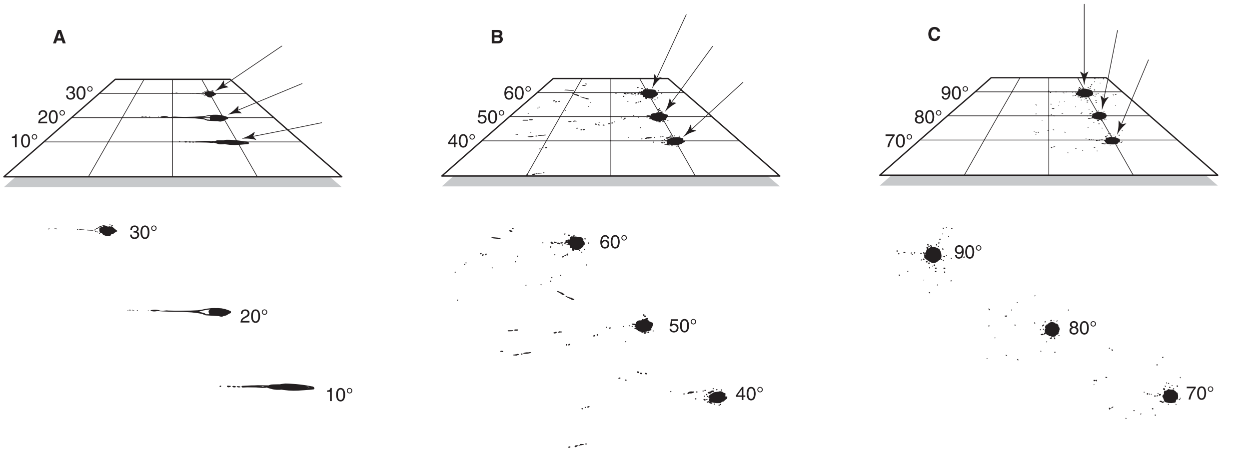


Figure 1 (**A**)–(**C**) The angles of impact of bloodstains against a target surface.

The tail is an important tool when reconstructing bloodstain patterns because it points in the direction the droplet was traveling in when it impacted on to the target surface. The shape of the bloodstain is also important in the reconstructive process. When the blood drop is deposited on to a surface that does not distort or alter its shape, the width-to-length ratio can be used to calculate the angle of impact for the bloodstain (**Fig. 5**). The most frequently used formula for this calculation is:

$$\text{Impact angle} = \text{arc sin (stain width} \div \text{stain length)}$$

The direction of travel for multiple bloodstains in a pattern that combined with the angle of impact determinations, is used to find the location of the blood source that was impacted to create the pattern (**Fig. 6**).

Surface effects on bloodstain appearance are very important. If a blood drop is deposited on a smooth, hard, nonporous surface, the shape of the bloodstain as well as the width-to-length ratio can be used for pattern reconstruction and interpretation. If the surface has characteristics which distort the appearance of the bloodstain or destroy its shape, then the stain will not be useful for a pattern reconstruction. Absorbent surfaces can also affect the shape and size of the bloodstain. These surfaces may alter the width-to-length ratio of the bloodstain and produce an inaccurate angle of impact determination; this data

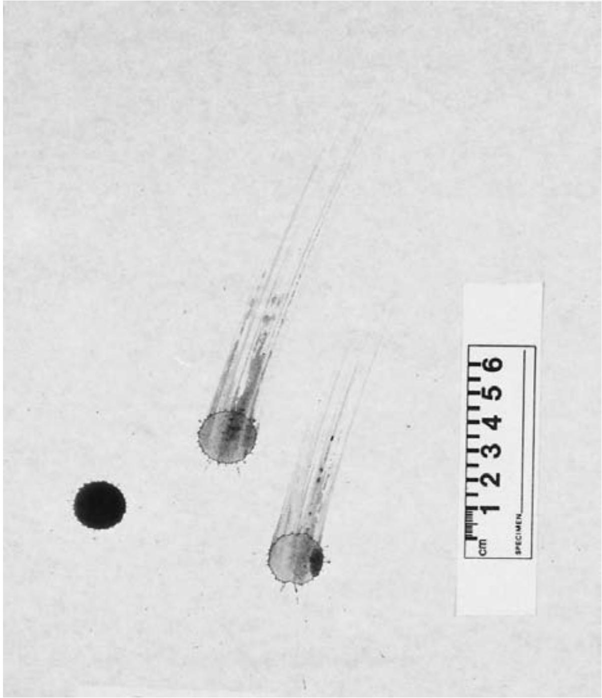


Figure 2 The shape of a drop of blood that impacted the target surface from an angle of 90°.

Figure 3 The shape of a drop of blood that impacted the target surface from an angle of 50°.

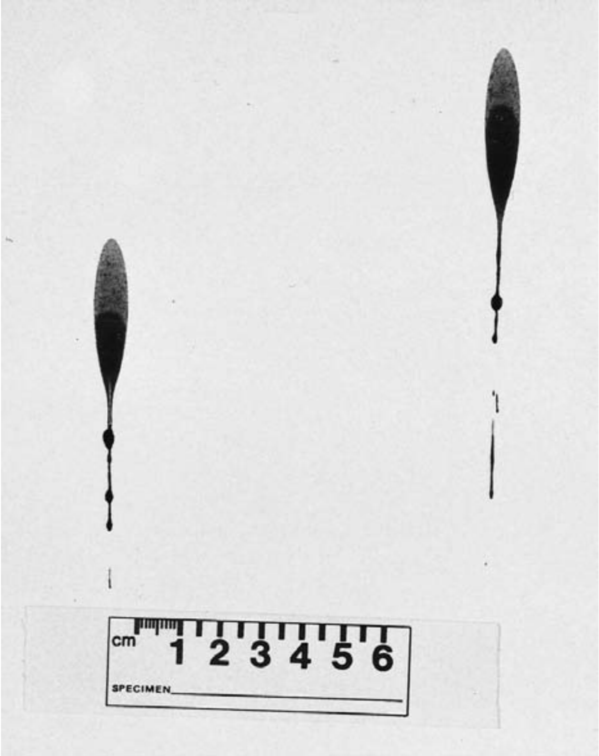


Figure 4 The shape of drops of blood that impacted the target surface from an angle of 12°.

could prevent an accurate pattern reconstruction for locating the pattern blood source. When a drop of blood strikes a surface that is very rough, it will cause the blood drop to break up. Determining the angle of impact for these bloodstains cannot be done because it is impossible to measure the width and length of the bloodstain.

Impact Bloodstain Patterns

Impact bloodstain patterns result when static pools of blood are hit. The energy of the impact is transferred to the blood, causing it to break up into droplets which are propelled through the air. If the droplets of blood hit an intermediate surface before the energy which put the blood in flight can dissipate to a level were gravity takes over, distinctive blood patterns will result (**Fig. 7**). The appearance of these blood patterns is affected by the forcefulness of the impact, the volume of blood impacted, the surface characteristics of the blood source, the characteristics of the surface the blood is deposited on, and what is impacting into the blood to put it in flight (**Fig. 8**). Impact bloodstain patterns will have round bloodstains in the portion of the pattern which is closest to the blood source. As the distance between the blood source and the bloodstains increase, the length of the stains increase and the width decreases. These bloodstains will also show the direction the blood drop was traveling when it impacted the target surface. These stain characteristics make it possible to reconstruct these patterns and determine the location of the blood that was impacted. The number of stains in the pattern, as well as the size of the stains, depends on the forcefulness of the impact. The majority of the small stains will be close to the blood source, while the larger stains tend to be longer distances away from the blood source. The number of stains in the pattern is dependent on how hard the blood was hit. Patterns which result from low force impacts have a low number of stains, the majority of which are large and have diameters greater than 3 mm. The patterns will cover a smaller surface area than patterns resulting from high force impacts. High force impact patterns have more bloodstains in the pattern and the majority of the stains that are close to the blood source will have diameters of 1 mm or less. The size and shape of the bloodstains and their distribution on the target surface make it possible to interpret bloodstain patterns and determine, within a reasonable degree of scientific certainty, the level of force used to impact the blood source.

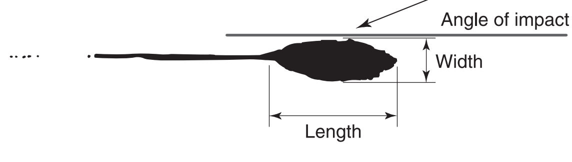


Figure 5 Measurement of the blood drop to determine angle of impact.

Cast-off Bloodstain Patterns

When liquid blood coats the surface of an object that is being swung, the blood will be thrown off of the object's surface due to the centrifugal force of the swing. If the blood is deposited on a target surface, the resulting pattern is called a cast-off pattern (**Fig. 9**). The bloodstains in cast-off patterns line up with one another, giving the pattern a linear appearance. The bloodstains in the pattern that are closest to the start of the swinging motion will be round, while those at the terminal end of the pattern (cast-off at the end of the swinging motion) will be oval. If the pattern is deposited on multiple surfaces that are oriented differently to one another, such as a corner where two walls meet, the shape of the stains in the pattern will be round wherever the arc of the swing was perpendicular to the target surfaces. The width of the cast-off pattern is a factor of the width of the surface the blood is forced from.

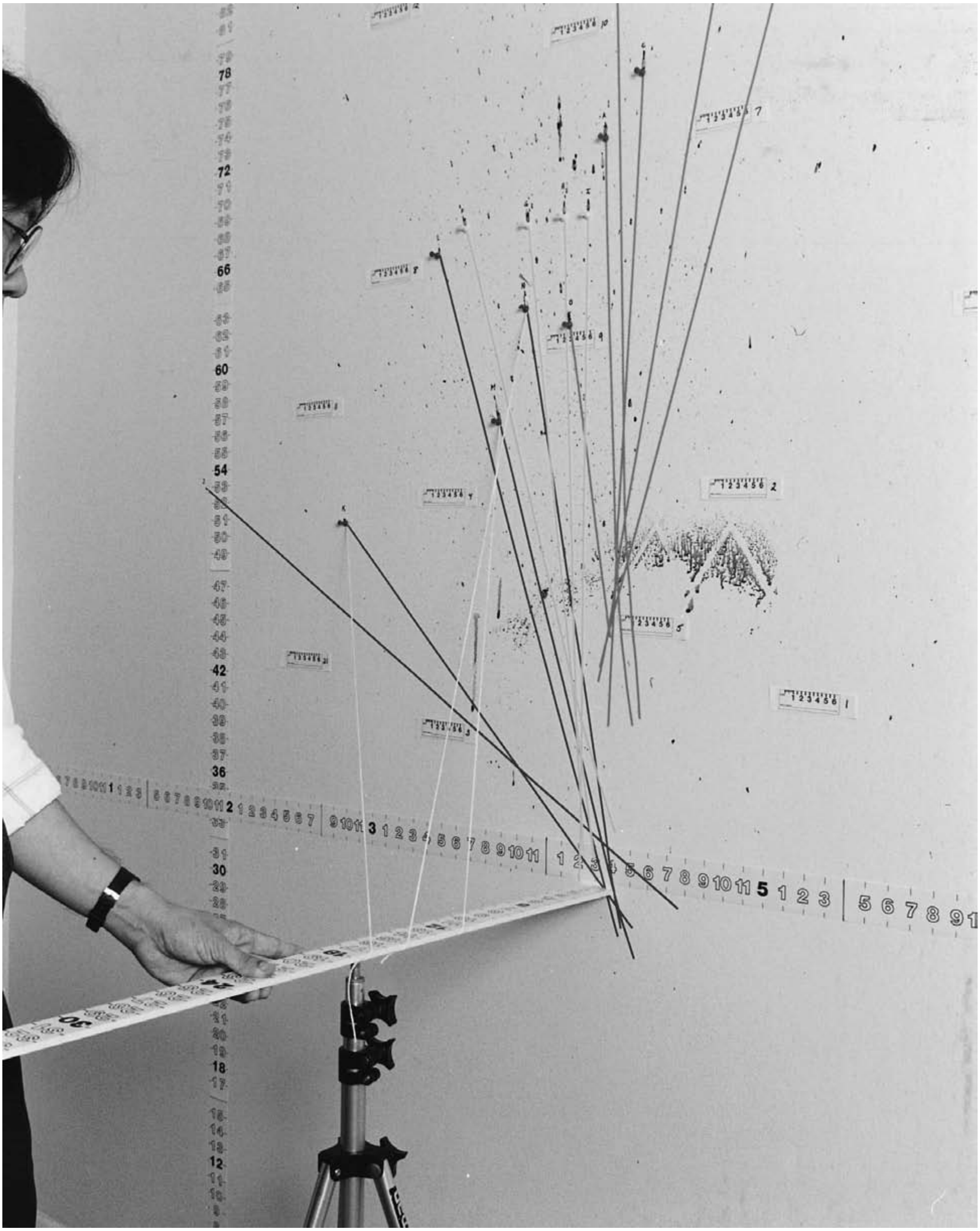


Figure 6 (see color plate 47) Bloodstain pattern reconstruction.

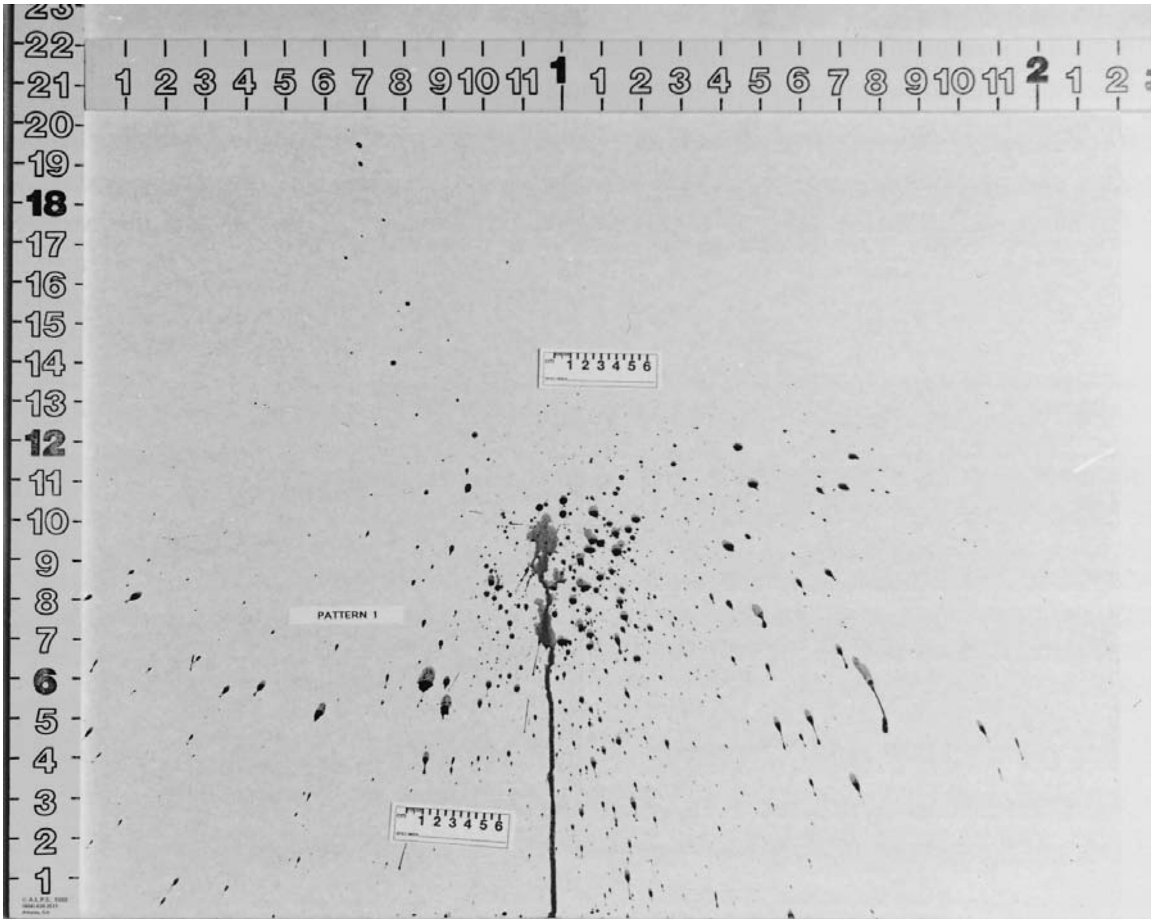


Figure 7 A low force impact pattern.

The general appearance of the pattern can be affected by the surface characteristics of the object, as well as by the motion and speed of the swing. Cast-off patterns usually occur only during the back swing away from the blood source. When a blood-coated object is swung forcefully, most of the blood will be cast-off during the back swing. The effects of inertia as the swing switches directions will force any remaining blood to be cast-off before the forward motion begins. Hitting a blood source several times can produce overlapping cast-off patterns. Since it is improbable that each swing will travel along the same arc, the overlapping patterns can be counted to indicate the number of hits after blood flow began. Cast-off patterns may exhibit a slight curvature along the length of the pattern. Under some conditions, this curvature can be used to indicate if the swing was left- or right-handed. Because cast-off patterns are easy to misinterpret, it is important to be cautious when evaluating these patterns for the number of hits or whether the swing was right- or left-handed.

Projected Bloodstain Patterns

Projected bloodstain patterns result when the blood is propelled forward against a target surface (**Fig. 10**). In this situation, the force that put the blood in flight is traveling in the same direction as the blood pushing it from behind. The action is similar to that of water squirted from a squirt gun. The most common cause of these patterns is arterial injury. Blood flows through our bodies in a closed system of veins and arteries. When an artery is damaged, blood will be forced out of the injury with every beat of the person's heart. The blood exits the artery under high pressure. If the blood strikes a target surface while the pressure or force is still elevated, the resulting pattern will have a very distinctive appearance. The center of the pattern will usually be one large bloodstain, or several large bloodstains. As the blood impacts into the target surface, the blood is broken up into hundreds of droplets which are splashed outward at an acute angle to the target surface, creating hundreds of secondary bloodstains that are very long and spindly in

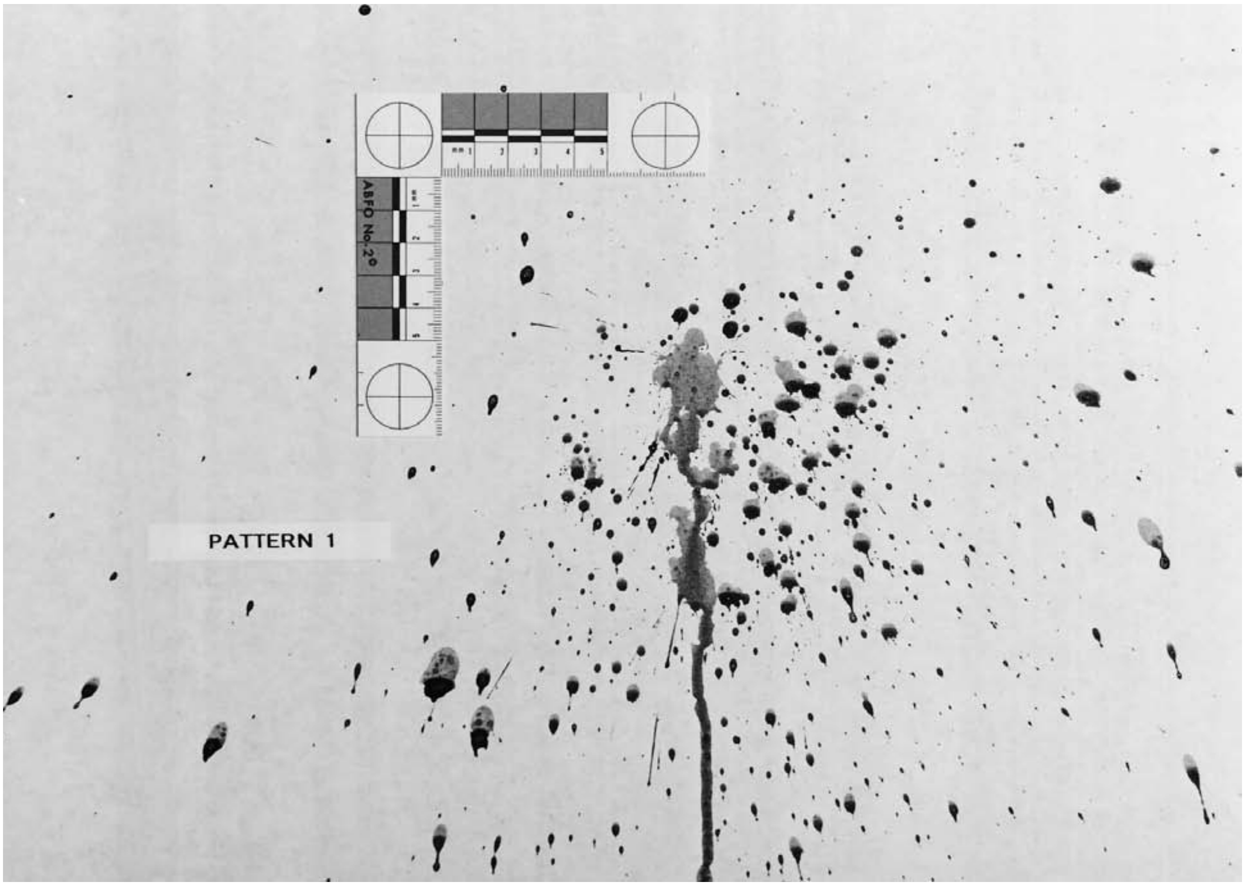


Figure 8 The characteristics of the bloodstain droplets near the center of the impact pattern.

appearance. With each successive beat of the heart, more blood will be pumped out of the injured artery, but the pressure under which it is forced out decreases due to blood loss. The patterns that occur later will not hit the target surface as forcefully, and the secondary spatter bloodstains lose the spindly appearance. These lower force projected patterns can also occur when the projected blood travels an extended distance from the blood source to the target surface. As the blood flies through the air, the energy that propels it forward will decrease until the speed the blood is traveling at has slowed down to the pull of the Earth's gravity.

Large Volumes of Falling Blood Patterns

Patterns that result from large volumes of falling blood will usually have a large central stain where the blood first impacts the target surface, with numerous secondary spatter stains radiating out from the impact site (**Fig. 11**). Bloodstain patterns that result from large volumes of falling blood will vary in appearance due to the effect of the distance the blood falls to the target surface. Large volumes of falling blood fall due to gravity. As blood falls, its velocity increases. The velocity of the falling blood will continue to increase until it reaches terminal velocity, $9.8\,\mathrm{m\,s^{-2}}$, the downward pull of gravity. Studies done by MacDonell on 0.05 ml drops of falling blood showed that terminal velocity would occur when the drop falls 25.1 feet (7.7 m). There is a rapid acceleration of the falling blood in the first 48 inches (1.2 m) of the fall, after which the acceleration is gradual until terminal velocity is reached. Blood that falls a short distance impacts the target surface at a low velocity, causing a less forceful splash and less secondary spatter. As velocity increases, the blood will impact the target surface more forcefully, the secondary spatter will be more numerous, and the bloodstains may start to become long and spindly. It is possible for bloodstain patterns from falling blood to be confused with projected bloodstain patterns. As the blood falls, its velocity increases until it impacts a target surface or reaches terminal velocity. The velocity of projected blood decreases until it impacts a surface or it reaches terminal velocity. If the distance

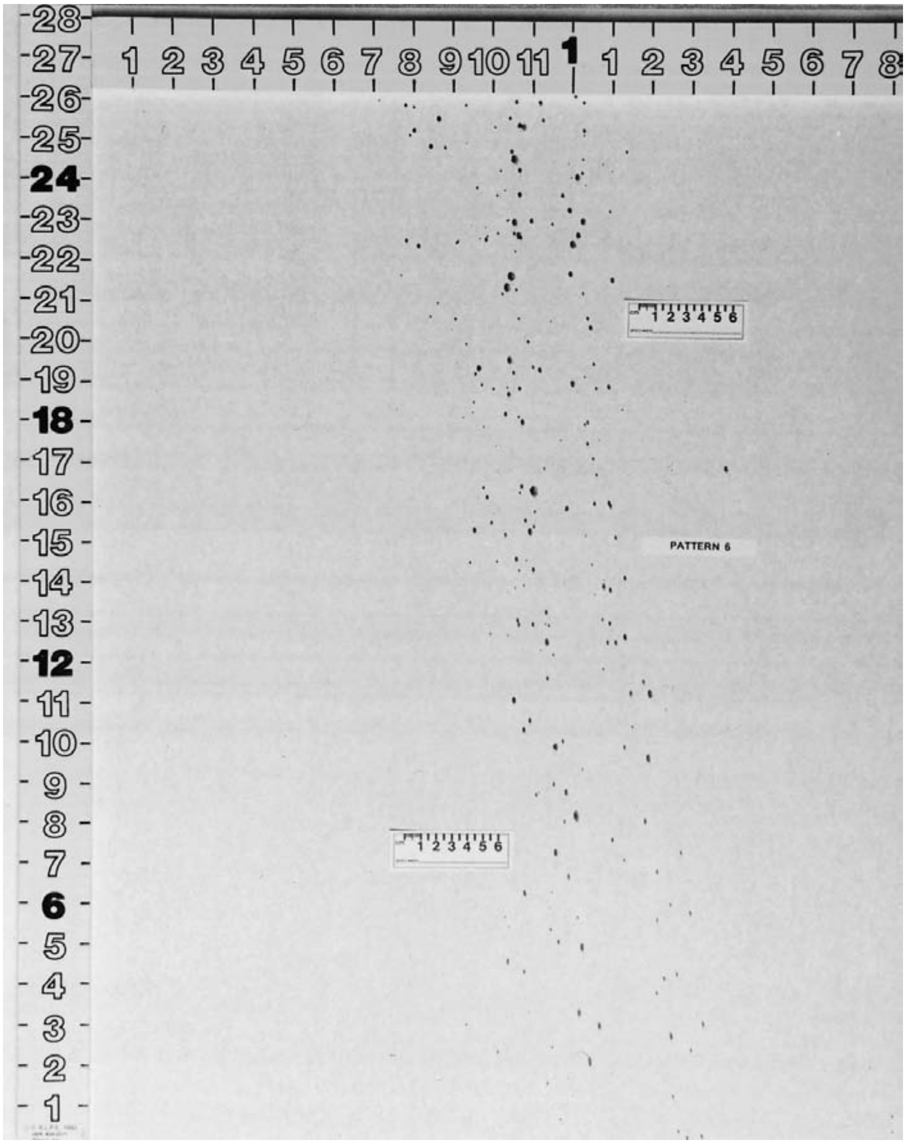


Figure 9 A cast-off bloodstain pattern.

traveled by the blood in these two situations is long enough, both volumes of blood will be traveling at the same velocity when they impact the target surface and the resulting patterns will appear similar.

Contact Bloodstain Patterns

Contact or transfer patterns happen when a bloody object comes into contact with an unstained surface. The blood on the object can be transferred to the surface, leaving a pattern. Contact patterns can be nondescript, lacking any characteristics of the origin of the pattern, or the pattern may exhibit some of the object's characteristics (**Fig. 12**). These characteristics can be compared to the object for the purpose of identification. The most common occurrence of this situation is bloody latent prints. These patterns are very important because they can be used to identify an individual and place him or her at the crime scene. They are also of importance when the latent print belongs to the subject and the blood the print is made from belongs to the victim. This situation indicates that there was some type of interaction between the subject and the victim after blood flow began. This type of comparison can also be done using bloody impressions from shoes, bare feet or any object that has individualized characteristics. Some contact patterns exhibit class characteristics, such as the shape of

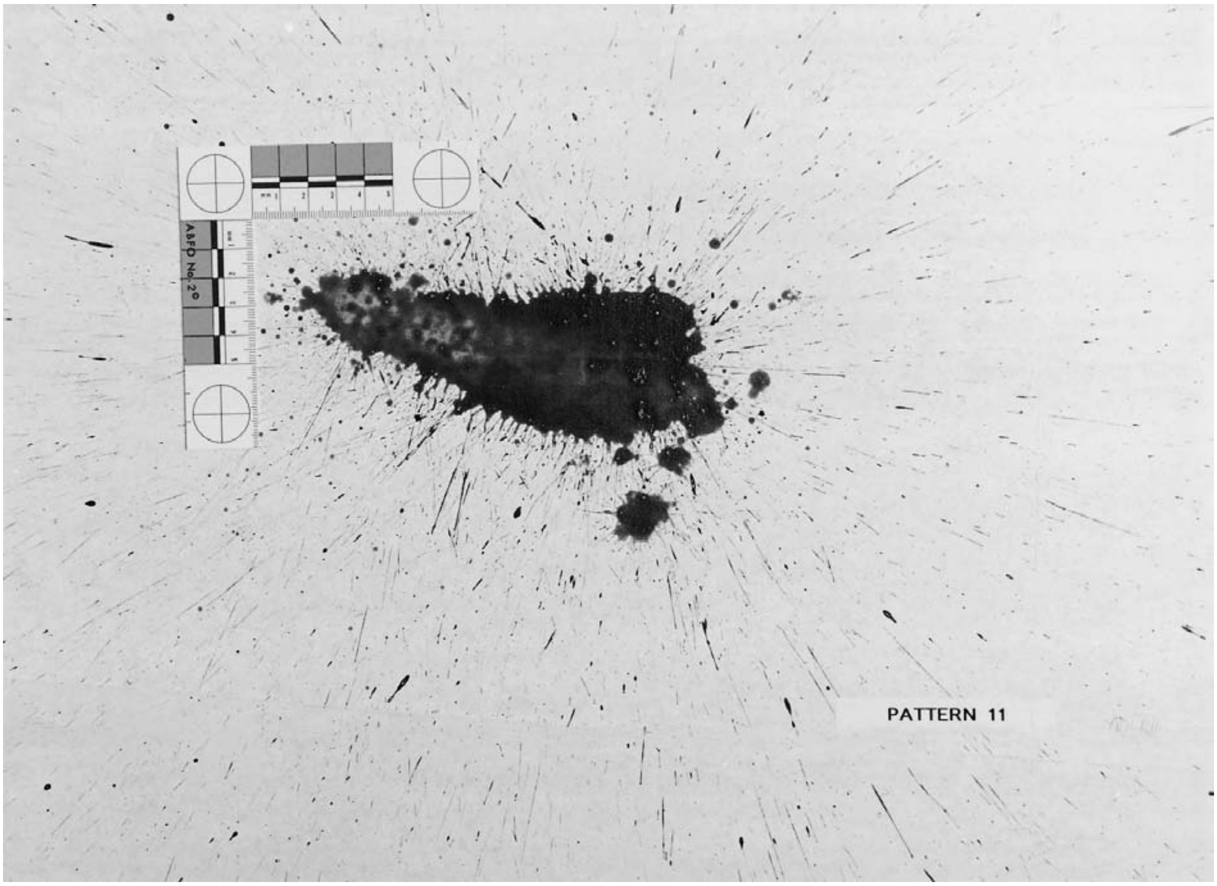


Figure 10 A bloodstain pattern that resulted from blood being projected at the target surface.

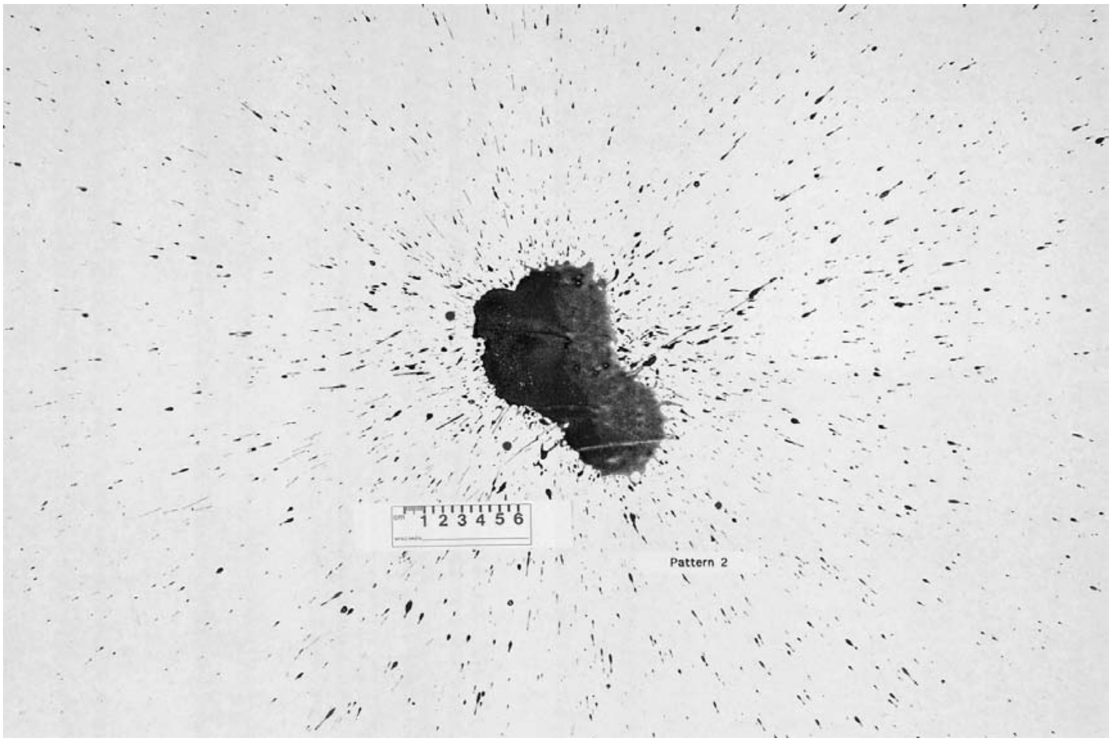


Figure 11 A bloodstain pattern that resulted from 5 ml of blood falling on to the target surface from a height of 1 m.

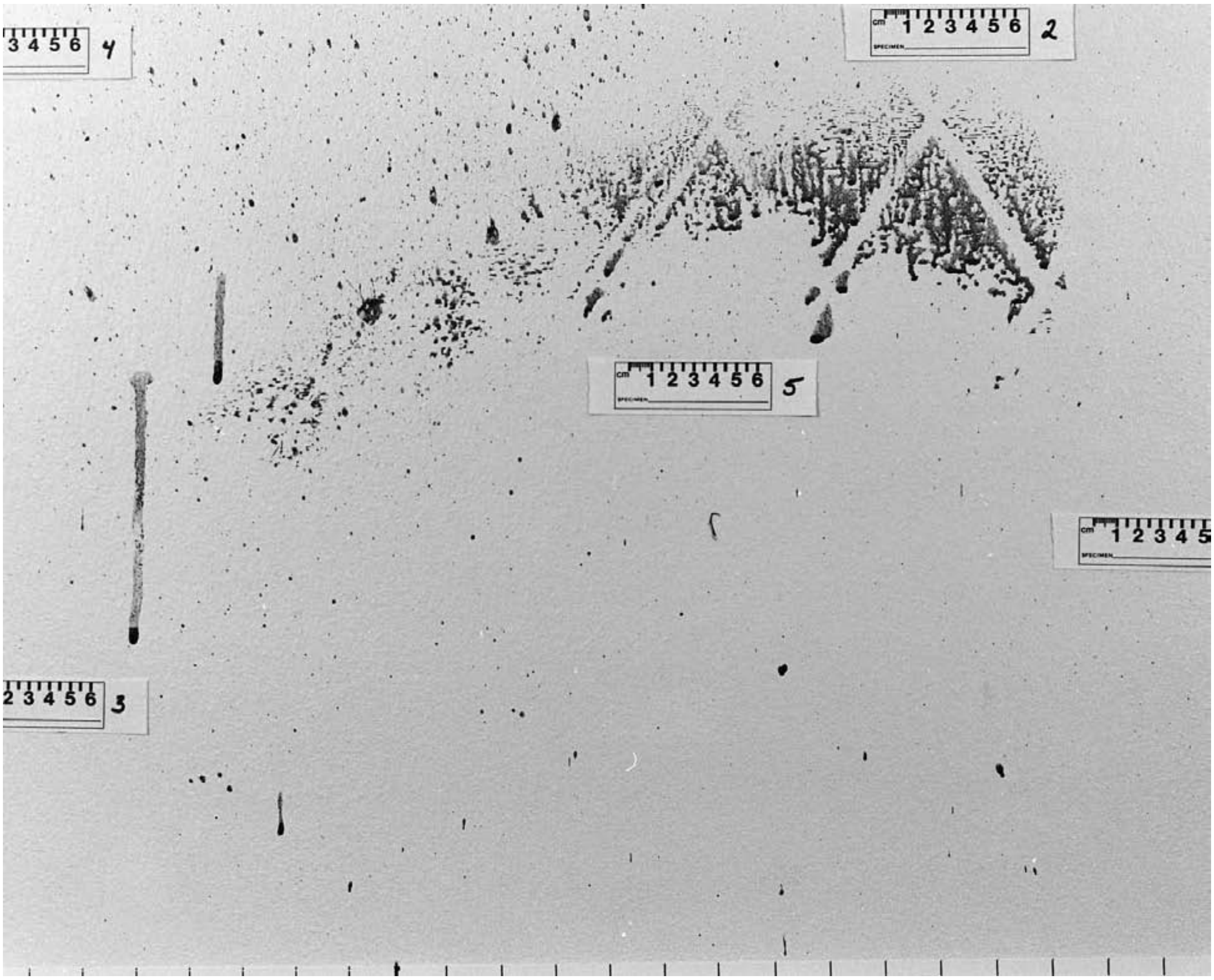


Figure 12 A transfer pattern that resulted from a blood-saturated mattress being leaned against a wall.

the object. These patterns can be useful for identification of the object type. If the object is wrapped while bloody, or the bloody object comes in contact with another surface, it may transfer blood, creating a pattern in its shape. Even though a pattern may lack details that can identify a specific object, the pattern is useful for investigative purposes.

Blood Trails

When blood drips from an open wound or from a bloody object as it moves horizontally, a blood trail will occur. Blood trails are characterized by large blood drops in a linear pattern that has been deposited on a horizontal surface while dripping from the blood source. This linear characteristic is due to the horizontal movement of the blood source above the target surface while the blood drips due to the pull of gravity (**Fig. 13**). The shape of the bloodstains and the degree of secondary spatter associated with the stains will be affected by the distance the drops fall and the speed of horizontal movement for the source of the dripping blood. These patterns can provide information about movements during and after bloodshed. For example, stains in a blood trail may show the direction of the blood source's horizontal motion. If the direction of travel can be determined, the trail can be followed. This is extremely useful when an injured individual leaves the crime scene. If nothing is done to obstruct the flow of blood, the trail may be followed to determine the location of the individual. Trails can also provide information concerning the sequence of events at a crime scene or assist in locating important evidence. In addition, trails can be characterized by drag marks rather than droplets of blood. When a blood-coated object is dragged over a surface or the object is dragged through blood on a horizontal surface, a trail of smeared blood will result. The direction of the horizontal motion can be determined by examining the fine detail of the pattern.

Figure 13 A blood trail.

Bloodstain Drying Times

Bloodstain pattern evidence drying times are affected by a number of environmental and physical factors. Blood volume is the main physical factor affecting the length of time required for the stains to dry. Large volumes of blood will dry more slowly than small volumes of blood when exposed to identical environmental assaults.

Environmental factors such as temperature, humidity, surface characteristics and airflow will affect the speed at which the bloodstain patterns dry. Warm temperatures will facilitate drying and accelerate the speed at which the blood dries. Cool or very cold temperatures will usually retard bloodstain drying times. Humidity also affects bloodstain drying times. When equal volumes of liquid blood are deposited in environments with contrasting humidities, the blood that is in the high humidity environment will dry more slowly than the blood in the low humidity environment. Drying occurs as the water in the stain evaporates into the surrounding air. In high humidity environments the air is saturated with water, retarding evaporation and increasing the length of time it takes for the stain to dry. The characteristics of the surface the blood is deposited on will affect bloodstain drying times. If the surface is one that protects the blood by limiting the amount of the bloodstain's surface that is exposed to the environment, drying of the stain will be retarded and result in longer drying times. When the bloodstain is deposited on a surface that maximizes the amount of the stain's surface exposed to the environment, the drying times will be shortened. Environments that allow a good airflow across the exposed surfaces of the bloodstain decrease the length of time required for the stain to dry.

Documentation of Bloodstain Pattern Evidence

Documentation and reconstruction of the bloodstain pattern evidence is a very important aspect of this forensic analysis. When the evidence and reconstruction are poorly documented, the conclusions can become very subjective. There are many different techniques that are currently in use for the documentation and reconstruction of bloodstain pattern evidence. Some bloodstain pattern analysts use computer programs that evaluate data from the bloodstain patterns and crime scene and determine the blood source locations. Many bloodstain pattern analysts do a physical evaluation and reconstruction of the bloodstain patterns at the crime scene, while others may work from the crime scene notes, sketches, and photographs.

Thorough and complete notes should accompany documentation and reconstruction of bloodstain pattern evidence. The notes should contain all observations and data that the analyst used to evaluate the bloodstain evidence. Contained in the notes may be police reports, crime scene reports, medical examiner reports and other forensic evaluations that may have been reviewed prior to the development of the analyst's conclusions.

The photographic record should show a quality set of pictures illustrating the bloodstain patterns that were evaluated and the reconstructions that were done. Before starting analysis or reconstruction, the patterns should be photographed and sketched. The camera should be positioned perpendicular to the surface the patterns are deposited on. This limits distortion in the pictures and gives a truer representation of the appearance of the bloodstain patterns. It may be necessary to use artificial lighting to illuminate the bloodstain patterns being photographed. The lights should be positioned to provide even lighting across the entire area of interest. This will help avoid creating areas which are overlit (too bright) or underlit (too dark). It is recommend that a handheld photographic light meter be used to assist in positioning the lights. After the bloodstain pattern evidence has been photographed in an unaltered state, measurement scales should be positioned around the pattern. Every picture should contain measurement scales that show the location of the pattern in relation to two fixed points. The measurement scales are also important in demonstrating bloodstain size data that are used in the bloodstain interpretation process. The bloodstain patterns should be rephotographed, showing the patterns with the measurement scales in place. Both of the aforementioned series of photographs should start with a general overview of the whole pattern(s) and then proceed to more specific pictures that show the detail of the blood droplets in the patterns. After the patterns have been documented in photographs, sketches and notes, the reconstruction may be performed. Each step of the pattern reconstruction should be photographed. Notes and sketches recording the data obtained should accompany the photographs from the reconstruction. Upon completion of this process, samples of the blood should be collected and submitted to a crime laboratory for analysis. Analysis of the blood samples may assist in the final interpretation of the patterns by helping the analyst connect the patterns to the individuals involved in the blood-spattering event.

See also: **Serology**: Blood Identification.

Further Reading

Bevel T and Gardner RM (1998) *Bloodstain Pattern Analysis. With an Introduction to Crime Scene Reconstruction*. Boca Raton, FL: CRC Press.

Eckert WG and James SH (1989) *Interpretation of Bloodstain Evidence at Crime Scenes*. New York: Elsevier.

James SH (ed.) (1998) *Scientific and Legal Applications of Bloodstain Pattern Interpretation*. Boca Raton, FL: CRC Press.

James SH and Eckert WG (1998) *Interpretation of Bloodstain Evidence at Crime Scenes*. Boca Raton, FL: CRC Press.

Laber TL and Epstein BP (1983) *Experiments and Practical Exercises in Bloodstain Pattern Analysis*. St Paul, MN: Minnesota Bureau of Criminal Apprehension Forensic Science Laboratory.

MacDonell HL (1982) *Bloodstain Pattern Interpretation*. Corning, NY: Laboratory of Forensic Science.

MacDonell HL (1993) *Bloodstain Patterns*. Corning, NY: Laboratory of Forensic Science.

MacDonell HL and Bialousa L (1971) *Flight Characteristics and Stain Patterns of Human Blood*. Washington, DC: United States Department of Justice, Law Enforcement Assistance Administration.

Wolson TL (1995) Documentation of bloodstain pattern evidence. *Journal of Forensic Identification* **45**:396–408.

Sex Determination *see* **Anthropology**: Morphological Age Estimation.

Sexual Assault *see* **Autoerotic Death**. **Clinical Forensic Medicine**: Sexual Assault and Semen Persistence.

Sharp Injury *see* **Causes of Death**: Sharp Injury.

Skeletal Analysis *see* **Anthropology**: Animal Effects on Human Remains; Skeletal Analysis.

Spectroscopy *see* **Analytical Techniques**: Hyphenated Chromatographic-Spectroscopic Techniques; Spectroscopic Techniques; Spectroscopy: Basic Principles.

STALKING

J Hargreaves, University of Liverpool, Liverpool, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0743

Introduction

Stalking has been called a crime of the 1990s. The crime was legally created in 1990, when California passed the first state antistalking law. This legislation was drafted in the wake of five unrelated murders of women who were previously stalked. One of the victims was television actress Rebecca Schaeffer, who was shot and killed, on 18 July 1989, by Robert Bardo, a fan who claimed he had been obsessed with her for 2 years. Media coverage, and pressure from the Screen of Actors' guild, was influential in helping to convince California's legislature to enact a criminal antistalking law. In the following 3 years every state in the country had produced similar legislation. Stalking was defined thus:

> Any person who wilfully, maliciously and repeatedly follows or harasses another person, and who makes credible threats with the intent to place that person in reasonable fear of death or great bodily injury, or to place that person in reasonable fear of the death or great bodily injury of his or her immediate family is guilty of the crime of stalking ... 12022.7 (Cal. Penal Code 646.9)

The law was amended in 1994: now 'credible threat' need only place the victim 'in reasonable fear for his or her safety'.

In Britain, stalking was legally recognized by the introduction of the Protection from Harassment Act, 1997.

Legal Issues Arising From Legislation

Three key elements of stalking, arising from legislation, surfaced to define contextually the circumstances in which behaviors would amount to stalking: repeated pattern; threats made or implied; and intent and effect.

Repeated pattern

For a person to be charged with stalking, he or she must engage in a course of conduct where specific acts of harassment, or following, occur on more than one distinct occasion. A pattern of behaviour is looked for to establish continuity of purpose. The crime is not an isolated incident, but rather a series of acts taken together. In its broadest sense, legislation aimed to cover a host of activities that might fall under the umbrella of 'unconsented contact'. Unconsented contact, according to one writer, refers to 'any contact with another individual without that individual's consent, or in disregard that the contact be avoided or discontinued'. This definition of unconsented contact makes legally enforceable the traditional idea of privacy as 'the right to be let alone'. Typically, the activities that constitute a course of conduct must occur within a specified time period.

Threats made or implied

Antistalking law requires, at a minimum, that the target *feel* threatened by the actions of the stalker. Such threats do not have to be explicit, as long as the stalker's other actions create a threatening climate for the target. Since stalking criminalizes what might otherwise be considered legitimate behavior, based upon the fact that the behavior induces fear, the level of fear induced in a stalking victim is a crucial element of the stalking offense. Acts that involve annoyance or emotional distress fall below the fear of bodily injury or death that is necessary to found an offense.

Intent and effect

Most legislation requires that the stalker has the criminal intent to cause fear in the victim. The course of conduct must be 'willful' and 'purposeful', 'intentional' or 'knowing'. In addition to requiring that a defendant had an intent to bring about a certain result, most states require that the stalker intended to cause fear or emotional distress in the victim.

Stalking offenders often suffer under a delusion that the victim actually loves them. Or they believe that all they need do is demonstrate determination in their pursuit of the target for the target to love them. The stalker may not actually intend to cause fear; none the less, as long as the defendants know, or should know, that their actions cause fear and distress, the alleged behavior amounts to stalking.

Stalking: A Paradoxical and Indeterminate Offense

Specification of 'stalking' as a recognized crime did not happen all at once. The manner in which it evolved has resulted in an offense that is diverse, in respect of the behaviors it seeks to prohibit and in the characteristics of those likely to offend. The difficulty has arisen because antistalking legislation was drafted to combat the stalking of 'celebrities'. It was drafted in broad form to encompass the many differing ways by which 'fans' make contact with their victim. The wording was such that it has subsequently been used as a source of legal redress for two other classes of complainant: those enduring 'obsessional harassment' (sometimes referred to as psychological rape), and those experiencing 'domestic violence'.

It has been suggested that the amalgamation of behaviors associated with obsessional harassment and domestic violence with those of celebrity stalking has led to domain expansion with regard to four aspects of the criminal offense of stalking:

1. the motivation of the offender;
2. the nature of the prior relationship between the offender and the target;
3. the types of behaviors associated with stalking;
4. the potential risk to the target.

Variability of offense and offender characteristics

Stalking has been called 'a paradoxical crime of the nineties' and 'an uncanny phenomenon'. The following aspects of stalking demonstrate the paradoxical and indeterminate nature of the offence:

- Motivation for the offender include such things as obsessional attraction of a media figure, love sickness, rejection rage and separation anxiety.
- The prior relationship between the offender and victim varies along a continuum of familiarity from that of stranger to acquaintance and ex-intimate partner or lover.
- Behaviors are varied and include many noncriminal activities. It is the repetitive and nonconsensual nature of the behaviors that mark them out as stalking. The effect of the behavior upon the victim is also material in elevating noncriminal behavior to the status of stalking.
- The mode of communication varies from physically close, face-to-face encounters, to remote, physically avoiding contacts.
- The tone of communications ranges, emotionally, from nominally affectionate to overtly aggressive.
- Potential risk to the target is unclear. Offense behavior has been described as essentially nonviolent, *and* as random violence; as potentially dangerous and likely to deteriorate; and possibly as a precursor to murder and rape.
- It has been noted that stalking has no fixed crime scene. The offender operates in and around the personal life space of the target, entering and leaving the target's personal domain at will.
- Unlike most other offenses, stalking has no time locus. Typically, there is no clearly definable beginning or end. Offense behaviors may be transitory or they may persist for years. Unwanted attention may be continuous or punctuated by periods of quiescence.
- The mental health of stalkers is also a variable factor. Some offenders have recognizable mental illness, such as erotomania (delusional), borderline erotomania (nondelusional), schizophrenia or personality disorder.
- The offence of stalking is gender neutral. However, most studies and surveys have concluded that the perpetrators of stalking are predominantly male, and the victim is typically female.

Stalking Behavior

Stalking comprises a broad range of behaviors. These range from proximal, face-to-face contact, to distal communications by mail or telephone. The stalker may make repeated phone calls; mail a stream of unsolicited letters or send unwanted taxis or pizzas. Pursuit may be on foot or by car. The perpetrator may send gifts, cards, letters and sinister objects. Actions may include trespass and burglary and involve the theft of trophies, diaries, address books and valuables. The offender may camp outside the home or workplace of the victim and mount surveillance. The tone of communications range from nominally affectionate to threatening and abusive. The perpetrator may threaten violence and may go on to carry out the threats. Actions are generally directed at the victim but they may also be directed at the friends and relatives of the victim. Whatever form the behavior takes, offenders ceaselessly, persistently and monotonously attempt to insinuate themselves into the personal life of the victim. Stalking is an offence in which the rules governing social interaction are continuously broken. In all cases, stalking and unwanted pursuit demonstrate a disregard of interpersonal boundaries.

Prevalence of stalking

The first ever national survey of stalking and its impact, was undertaken by the National Institute of Justice Centers for Disease Control and Prevention (NIJCDC) (*Stalking in America*: findings from the national violence against women survey, 1998). This

body, part of the US Justice Department, reported the results of a nationally representative telephone survey of 8000 women and 8000 men in April 1998. Prior to this, there had been little empirical data on such things as the prevalence of stalking, and who stalked whom. The survey defined stalking as 'a course of conduct directed at a specific person that involves repeated visual or physical proximity, nonconsensual communication, or verbal, written or implied threats, or combination thereof, that would cause a reasonable person fear'. ('Repeated' meant two or more occasions.)

The survey found that 8% of women and 2% of men in the USA had been stalked at some time in their lives. Using the US census they estimated that 1 out of every 12 American women (8.2 million), and 1 in every 45 American men (2 million), has been stalked at some time. The survey also found that 1% of all women and 0.4% of all men had been victims of stalking in the preceding 12 months, giving an indication of the estimated number of persons stalked annually.

Age of victims when stalking started

The NIJCDC survey found that persons between the ages of 18 and 39 years were most at risk of being stalked. Seventy-four percent of stalking victims were in this age band and they represented nearly half of the population sampled. Similar results have been found by other researchers. Recently, in a forensic sample of 131 offenders, the age of 97 targets was known. Of these 97, 63% were between 20 and 39 years.

In the NIJCDC survey, the primary targets of stalkers were young adults aged 18–29 years. These accounted for 52% of the victims. In the 30–39 year age group, 22% were victims. In the above forensic sample, it was also found that young adults were more at risk, with 34% of targets falling between 20 and 29 years.

Who stalks whom?

Most victims of stalking in the NIJCDC survey were known to their stalker, confirming previous reports signifying that intimate and acquaintance stalking is a greater problem than the stalking of celebrities. The survey found only 23% of female victims (n = 650) and 36% of male victims (n = 179) were stalked by complete strangers. Fifty-nine percent of female victims, compared with 30% of male victims, were stalked by an intimate partner. In the forensic sample above, 77% (n = 131) of targets were stalked by someone they knew. Sixty-eight targets (51%) were stalked by ex-intimate partners who were husbands, cohabitees or noncohabitees. A further 25% were stalked by acquaintances. The remaining 30% of targets were stalked by complete strangers.

The NIJCDC survey reported that in intimate relationships there is a strong link between stalking and other forms of violence. Of women who were stalked by their current or former husband or cohabitee, 81% were physically assaulted. Of these, 31% were also sexually assaulted by their partner.

Ninety percent of stalking victims in the survey were stalked by one person. This concurs with other studies. Although stalking is a gender-neutral crime, women were found to be the primary victims of stalking and men the primary perpetrators. Seventy-eight percent of the surveyed victims were women and 22% were men, whereas 87% of the stalkers identified by victims were male.

In most cases of stalking, the target is of opposite gender to the perpetrator; however, cases of female homosexual erotomania have been reported. Cases of male homosexual erotomania have also been reported.

Motivation for Stalking

Many motives have been ascribed to stalkers. Perpetrators may be seeking justice or pardon. They may be motivated by a need for acknowledgment, respect or affection. Retaliation, revenge and punishment may be also be included. Offenders may be fixed and resolute in what drives them to pursue the target, but more often, the motivating forces will ebb and flow and change as stalking progresses. Motivational forces may be subject to the myriad mixture of often contradictory emotions being experienced by the offender. Initial feelings of love may be followed by jealousy, anger or separation anxiety. Stalking is embedded paradoxically in a range of contradictory emotions, such as love and hate, acknowledgment and denial, acceptance and rejection, attraction and repulsion. Motivation for stalking is often bound up in the prior relationship between the stalker and the target. Some individuals stalk to attain, or maintain, control of the victim. In other cases the stalking behavior is a result of inept attempts to initiate a new relationship. In some cases stalking is concerned with a realigning of the power in an existing relationship, and arises directly from a power shift in that relationship.

It has been said that stalking 'typically occurs where there is perceived to be an extraordinary, fundamental (potential or materialized) bonding between two people… It is an attempt to rescue the relationship from change or termination which are experienced as insufferable.'

A situational classification of stalkers has also been suggested; this includes obsessional fans, divorced or separated spouses, ex-lovers, rejected

suitors, neighbors, coworkers, classmates, gang members, former employees, disgruntled defendants, as well as complete strangers.

Domestic violence

It has been reported that when women are stalked by intimate partners, the stalking often occurs after the woman leaves the relationship. The NIJCDC survey observed that stalking occurred after the relationship ended in the case of 43% of respondents. They noted, however, that in 36% of cases stalking had occurred both before and after the relationship ended. Accordingly, it has been said that former intimate stalkers often 'have an emotional history of emotional dependence upon their partner that is severed when the relationship is terminated'. This leads to a jealous need to control the former partner. Also that in many cases 'he is so dependent on her that he would kill her rather than let her go and not be able to live without her'.

One writer has defined three stalking motivations of former intimates: the jealous lover, the violent husband, or the vengeful ex-husband. Stalking is, according to this definition, 'a type of domestic violence, which takes place once a battered woman breaks the abusive relationship and attempts to leave'. From this perspective it has been suggested that stalking should be viewed as an additional phase in the cycle of wife-battering. Stalking would be phase four. Three phases in wife-battering have been identified: the tension-building phase; the acute phase; and, finally, the tranquil, nonviolent, loving stage. The fourth, stalking, phase would begin 'when a battered women finally attempts to break the cycle of violence by terminating the abusive relationship'. It is at this time that 'her decision to leave is often met with escalated violence, in the form of stalking'. From this perspective 'marital separation cases which have been identified as stalking incidents merely constitute another phase in the domestic violence cycle ... phase four.'

Most writers do not accept, however, that stalking is merely an additional phase in the battered wife syndrome.

Celebrity stalking

In celebrity stalking, the victim is a famous person: a television performer, film star or other public figure. It has been noted that media reports typify the gender of the offender as male or female and the relationship between the stalker and victim as that of stranger. Motivation has been defined as that of obsessional attraction to a famous person. Actions include the sending of letters and gifts, attempted contact by telephone, visits to appearance venues, surveillance and stalking. The behavior of the offender has been linked in some cases to erotomania, a 'delusional (paranoid) disorder eroto type' (DSM-III-RY, *Diagnostic and Statistical Manual of Mental Disorders*, 3rd edn, revised). The stalkers' behavior has been characterized as unpredictable, potentially dangerous and likely to deteriorate over time. Research into celebrity stalking has focused on the characteristics of threatening and inappropriate letters to high-profile victims such as Hollywood celebrities, and members of Congress. The aim was to identify potential risk factors, or psychological indicators, associated with subsequent approach behaviors; however, the results have limited applicability to the stalking of noncelebrities.

Sociopathic stalking

According to one writer, 'two groups of criminals that are notably absent from the "stalking" literature are serial murderers and serial rapists, an absence that is all the more remarkable given that the behavioral activities proscribed by most stalking laws are familiar characteristics of such individuals'.

There are many accounts of serial murderers who reputedly stalked their victims before killing them. One such offender is serial murderer Ted Bundy, who started his serial raping and murdering when the woman he loved terminated their relationship. One writer has commented that this 'pattern of behavior is similar to those of former intimate stalkers except that his victim was not the former intimate herself'. The writer concluded from this that 'Ted Bundy's killings are an excellent example of the outcome of a sociopathic stalker's displacement.' The life histories of serial murderers have been likened to those of borderline erotomanics and former intimate stalkers.

Classification Systems and Typologies

The indeterminate nature of stalking legislation has resulted in constitutional difficulty. No single definition of stalking has emerged that satisfactorily accounts for the many differing behaviors observed. A wide range of classification systems have been employed to explain the phenomenon of stalking. The literature supports classification of stalker characteristics, target characteristics, prior relationship between the perpetrator and the victim, and location of stalking. More recently, research from one center has focused on classification of stalking actions.

Identifying the stalker

A number of classification systems have been proposed for stalkers. Mental health professionals identify three main categories of disorder that are generally attributed to stalkers of celebrities. These are: (1) delusional mental illness, such as erotomania and/or schizophrenia; (2) borderline erotomania, where no delusion is present but where there is an extreme attachment disorder; and (3) personality disorder.

Erotomania The stalker may suffer from de Clerambault's syndrome (erotomania). This is a delusional disorder, erotomanic subtype (American Psychiatric Association). Erotomania is the delusional belief that one is passionately loved by another. These individuals make great efforts to contact the object of their obsession, who is usually a person of higher socioeconomic class or status or an unattainable celebrity figure. It has been said that the erotomanic 'seeks to establish an intimate, even permanent, relationship with the object of fantasy.' Moreover, that 'this delusion usually focuses on idealized romance or spiritual union rather than sexual attraction'.

The term 'borderline erotomania' was used in a small group study to describe a sample of individuals who stalked their object but who did not show obvious signs of delusion. These individuals developed intense feelings for individuals who they knew did not reciprocate their feelings.

This definition of borderline erotomania does not seem to have been accepted by the psychiatric community – one influential opinion being that the term for this syndrome (nondelusional or borderline erotomania) is confusing and should not come into general use because 'nondelusional erotomania is a contradiction in terms, as erotomania's delusionary nature has, for the last 100 years, been considered its very essence. In addition, the word "borderline" has at least two meanings ....'

This term was seemingly rejected in a study comparing erotomanic to nonerotomanic stalkers. The study focused on behavioral patterns, potential risk to victims and means of intervention. Three distinct types of subject were identified, based on the quality of their obsessions: erotomanics, love obsessionals and simple obsessionals. Love obsessionals, like erotomanics, usually do not know the victim, except perhaps through the media. In this group, many hold the delusion of being loved by the object of pursuit, but this is only one of several delusions and psychiatric symptoms. With simple obsessionals, a prior relationship between the offender and the victim has existed. The relationships in this group varied in degree from customer, acquaintance, neighbor, professional, dating and lover.

It has been suggested that psychiatric classification of stalkers favors one point of view, that of the psychiatrist. From this perspective, 'stalking' is labeled the product of a sick mind and 'the value system that is reflected and reinforced is the psychiatric ideology'. However, in recent studies a number of mental health clinicians appear to have eschewed the term 'stalking' altogether, for clinical populations. Instead 'patients' are referred to as 'obsessional followers'. One of the reasons given for avoiding the term 'stalking' was 'to avoid mimicking its sensationalistic use by the popular media, and to reserve its proper use for the description of a defined criminal act ....' The term 'obsessional following' has been defined as 'a stalking behavior in which a person engages in an abnormal or long-term pattern of threat and harassment directed toward a specific individual'.

It has been suggested that, by defining the criminal offence of stalking as conduct of certain types of people, this establishes 'psychiatric soundness' as the social value protected by the criminal law of stalking. When an offence is determined by the doers rather than by its effect on a social value, law becomes psychiatry, defining people for what they are rather than for what they did. But unlike psychiatry, the law should be designed to evaluate people's actions and not their personalities.

Importantly, the NIJCDC survey reported that 'only 7% of the victims said they were stalked because their stalkers were mentally ill or abusing drugs or alcohol'. They went on to say that these 'results dispel the myth that most stalkers are psychotic or delusional'.

Identifying the victim

There have also been various attempts to classify the victims of stalkers. One classification system proposed the following categories of victim: television and film stars; those within the entertainment industry who are less recognized; executive level and supervisory personnel; and ordinary citizens with no distinguishing history.

Many studies have defined victims in terms of their relationship with the perpetrator. Most studies have categorized two groups, based on the level of intimacy prior to stalking. Classification in one study divided victims into strangers and former intimates; the latter included those victims who had had any sexual intimacy with the perpetrator prior to the stalking. The two classes of relationship proposed

by another study was domestic and nondomestic relationships: domestic included former boy/girl friend, family or household member, common-law relations and long-term acquaintances; nondomestic included those with no relationship with the victim and unknown stalkers. These two relationship classes were further divided into two types based primarily on the presence or absence of delusion. A recent forensic study has identified four types of relationship between the offender and the target: stranger, acquaintance, noncohabitee and cohabitee. A further study divided victims into seven groups: personal, professional, employment, media, acquaintance, no relationship and unknown. These groupings formed one axis of their classification system.

Having classified stalkers and victims, a number of studies have attempted to identify types of actions or behaviors that seem to be associated with these classes of offender. A reading of these studies suggests there is much overlap of behaviors across categories of stalkers.

The various classification systems and typologies make it difficult to compare the results of the various studies of stalking. It has been suggested that 'until reasonably clear classification of stalking *actions* can be derived, it is difficult to see how our understanding of stalking can progress'.

Identifying stalking actions

It can be seen from the above that previous studies have focused primarily on the classification of stalkers and stalking victims. Prior relationship between the stalker and target has been seen primarily as a demographic feature of the victim. Classification of stalking actions, on the other hand, has been confined to the distant contact behaviors of celebrity stalkers.

The most recent empirical analysis of stalking actions comprised 2636 stalking events in a forensic sample of 27 offenders. The results suggest that there were core actions commonly present in many stalking events. However, some stalking actions were found regularly to co-occur in particular stalking events. These behaviors were linked thematically according to the manner in which offenders interacted with the targets, and the way in which they used knowledge and information about the targets. This study identified four variations in stalking style based on: (1) the detached or attached interpersonal style of the offender; and (2) whether or not the offender's goal was exploring or exploiting knowledge and information about the target. This resulted in a fourfold classification of stalking actions: hunting, manipulating, oppressing and invading.

Interaction between stalking actions and prior relationship The relationship between the stalker and the target has been seen primarily as a demographic feature of the target. However, a recent study hypothesized variation in stalking actions based on the prior relationship between the stalker and the target. The researchers' sample comprised four relationship groups: strangers who had had no known contact with the target prior to stalking; acquaintances, comprising customers, friends and coworkers; noncohabitees, comprising former sexual intimates; and cohabitees (former sexual intimates who had shared a home). The study reported that prior relationship between the stalker and the target had a significant effect upon the style of stalking adopted by the offender.

Threats Made by the Stalker

The relationship between threats and subsequent violence is controversial. Clinical research has found that 3% of perpetrators who threaten homicide carried it out, while 4% committed suicide. Other research has found no relationship between approach behavior and threatening letters to Hollywood celebrities.

In the NIJCDC survey, less than half of all male and female victims were overtly threatened by their stalker. In spite of this, stalking victims reported that they were very frightened by the stalker's behavior or very fearful that the perpetrator would seriously harm or kill them. Fear was generated by the threatening climate created by the stalker's course of conduct, whether or not it was accompanied by direct threat.

When and Why Does Stalking Stop?

Ninety-two percent of respondents in the NIJCDC survey said that they were no longer being stalked. Approximately two-thirds of all stalking cases lasted a year or less, a quarter between 2 and 5 years, and around a tenth went on for more than 5 years. The average stalking case lasted for 1.8 years. Cases of stalking involving former intimates lasted twice as long (2.2 years) as nonintimate stalking, which lasted approximately 1.1 years.

Nineteen percent of respondents believed stalking had stopped because they had moved away. The stalker becoming involved in a new relationship was the reason stalking stopped in 19% of cases. A police warning was effective in stopping stalking in 15% of cases. The arrest of the stalker was less effective, resulting in cessation of stalking in only 9% of cases.

See also: **Identification/Individualization**: Overview and Meaning of ID. **Psychological Autopsies**.

Further Reading

Bernstein E (1993) Living under seige: do stalking laws protect domestic violence victims? *Cardozo Law Review* **15**:525–567.

Cahill P (1995) Unrequited love. In: *Pizer Medical Aspects of Human Sexuality*, **19**.

Dietze PE, Mathews D, Van Duyne C *et al.* (1991) Threatening and otherwise inappropriate letters to Hollywood celebrities. *Journal of Forensic Sciences* **36**:185–209.

Hargreaves JC (1999) *Differentiating Stalking: An interpersonal Model of Variation in the Stalking Process as a Function of Prior Stalker–Target Relationship*. PhD thesis. Liverpool: University of Liverpool.

Hargreaves JC and Canter DV (2000) Stalking: 'obsessional-harassment': an interpersonal model of variation in stalking styles. In: Canter D and Alison L (eds) *Profiling Rape and Murder*. Ashgate: Aldershot. (in press)

Harmon R, Rosner R and Owens H (1995) Obsessional harassment and erotomania in a criminal court population. *Journal of Forensic Sciences* **40**:188–196.

Lane JC (1992) Threat management issues. *Police Chief* **August**:27–31.

Lowney KS and Best J (1995) Stalking strangers and lovers: changing media typifications of a new crime problem. In: Best J (ed.) *Images of Issues: Typifying Contemporary Social Problems*. New York: Aldine De Gruytero.

McAnaney KG, Curliss LA and Abeyta-Price E (1993) From imprudence to crime: anti stalking laws. *68 Notre Dame Law Review* 819.

Meloy JR (1992) *Violent Attachments*. Northvale, NJ: Jason Aronson.

Meloy JR (1996) Stalking (obsessional following): a review of some preliminary studies. *Aggression and Violent Behaviour* **1**(2):147–162.

Wright JA, Burgess AG, Burgess AW *et al* (1996) A typology of interpersonal stalking. *Journal of Interpersonal Violence* **11**(4):487–502.

Zona MA, Sharma KK and Lane J (1993) A comparative study of erotomanic and obsessional subjects in a forensic sample. *Journal of Forensic Sciences* **38**:894–903.

Statistics *see* **Deoxyribonucleic Acid**: Significance.

Stature Estimation *see* **Anthropology**: Stature Estimation from the Skeleton.

STR *see* **Deoxyribonucleic Acid**: Polymerase Chain Reaction-Short Tandem Repeats.

Stress *see* **Anthropology**: Assessment of Occupational Stress.

Suspicious Deaths *see* **Crime-scene Investigation and Examination**: Suspicious Deaths.

T

Tachographs *see* **Accident Investigation**: Tachographs.

TIME SINCE DEATH

A Tracqui, Institut de Médecine Légale, Strasbourg, France

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0613

Introduction

Providing an estimation of the time elapsed since death is a request almost invariably aimed at forensic specialists summoned to the scene of a death. Despite the fact that there has been an extensive literature on this topic for more than a century, this determination still remains difficult, even for experienced pathologists, and must be undertaken with extreme caution. In particular, the dogmatic application of rules and formulas based upon single and isolated observations (e.g. rectal temperature of the body, extent of lividity, stage of putrefaction) is a guarantee of inaccuracy because of the numerous factors of cadaveric or environmental origin that can influence the 'normal' rate of postmortem changes. The longer the time since death, the more important the weight of these factors and the more difficult it will be to estimate the death time with some accuracy.

The purpose of this article is to present a brief overview – focused on advantages and limitations – of the different techniques that can be applied to human corpses in order to determine the death time. As shown in **Fig. 1**, most of them concern the so-called 'early postmortem period' (usually defined as the time between death and the appearance of generalized putrefaction).

Body Temperature

Models of body cooling

After death, body temperature progressively declines until it reaches the temperature of its surroundings. Although this phenomenon has been appreciated since prehistory, the first papers dealing with its potential interest to the field of forensic medicine date back to the mid-nineteenth century. In temperate climates this process generally takes about 8–12 h at the skin surface, but the core of the body is known to require two to three times as long to cool down.

The cooling of a body (also called 'algor mortis') is determined by physical rules and involves four different mechanisms of heat transfer: conduction, radiation, convection and evaporation. Their respective influence on the fall of body temperature after death strongly depends on environmental circumstances (temperature, hygroscopy, air agitation, etc.). According to a rule of thumb usually employed in the past, time since death could be roughly estimated by assuming a temperature loss of 1°C per hour during the first 24 h. However, the rate of cooling of a physical body is *never* a linear function of time: under experimental conditions this phenomenon is mainly determined by the difference in temperature between the body and its surroundings, and should be represented by an exponenential function of time (Newton's law). As shown by many measurements on corpses, the reality is still much more complex, owing to several factors like the irregular shape

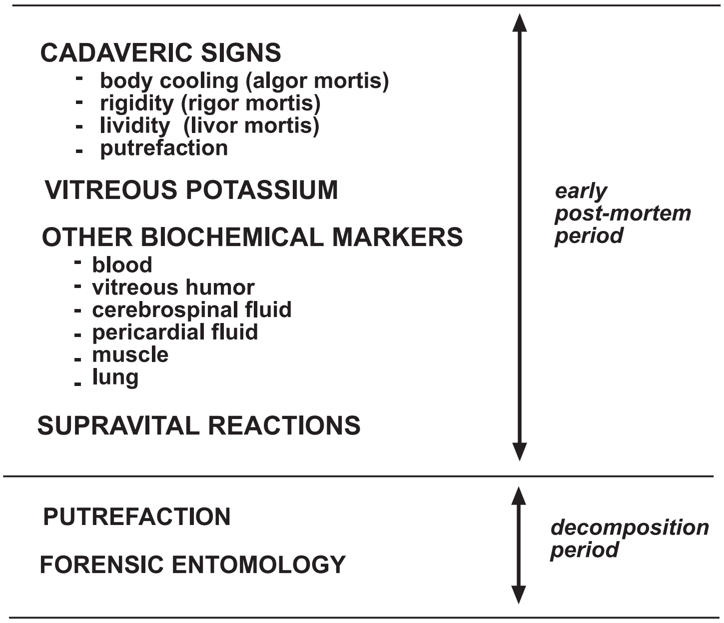


Figure 1 Common estimators of time since death.

and inhomogenous composition of the human body, or a moderate postmortem heat production resulting from bacterial metabolism. Finally, when temperature is plotted against time, the postmortem cooling of the human body is best represented by a sigmoid or biexponential curve that comprises three successive periods:

- *An initial phase*, in which the body temperature remains relatively even – the so-called 'temperature plateau' or 'lag period'. This period usually lasts 0.5–3 h, with important (and unpredictable) interindividual variations.
- *An intermediate phase*, in which temperature drops rapidly and almost linearly.
- *A terminal phase*, in which this drop progressively slows as the core temperature approaches that of the environment.

Temperature measurements

It is essential to measure the central temperature of the body as well as the environmental (air) temperature at the scene of death. Both must be taken at the same time and with the same instrument, and the time of measurement must be carefully noted. On the several anatomical sites that have been proposed for temperature determination, the rectum probably ranks first for convenience and reliability; however, this may be debatable if a sexual assault is obvious or suspected. The tympanic determination represents an interesting alternative for some authors. Endobuccal or external determinations (at the axilla or groin, or on the abdominal skin) must be avoided. Medical thermometers should not be used for rectal measurements because they are too short and their range of temperature is too narrow. The best instruments are thermocouple, electronic devices with digital display, which may usually be linked to a computer and/or printer. The thermometer must be equipped with a long, rigid probe that has to be inserted at least 10 cm into the rectum.

Advantages and limitations

If measured under appropriate conditions, body temperature may be held as one of the best estimators of the time since death during the first 24 h; however, it presents some limitations that should always be borne in mind:

- The technique is considered accurate only during the intermediate, pseudolinear phase of the sigmoid cooling curve. Otherwise, temperature measurements are not useful and may lead to inconsistent estimations during the very early postmortem period – the temperature plateau – or when the body temperature closely approaches that of the surroundings.
- All determinations of the time of death by the body cooling method assume that the body temperature at the time of death was within normal limits,

37.2 ± 0.4°C. Hypothermia or hyperthermia preceding death may result in mistakes and thus should be systematically searched for if antemortem data are available.

- Cooling equations also assume that the environmental temperature has remained constant over the postmortem interval. This may be the case when the death occurs in air-conditioned or heated buildings; however, problems may arise with corpses found outdoors, owing to the circadian changes of ambient temperature.

Furthermore there are a number of complicating factors, of endogenous (cadaveric) or exogenous (environmental) origin, that may strongly affect the rate of postmortem cooling. These include:

- *Body mass* Heat loss is slower in obese subjects (owing to greater body mass and because fat may act as an insulator), whereas it is more rapid in slender ones.
- *Movement of air* Any air movement enhances heat losses by convection, thus accelerating the fall in temperature. At outdoor scenes of death it is therefore important to note if the weather is windy, and in buildings sources of air currents (open doors/windows) should be looked for.
- *Humidity* High atmospheric humidity will increase heat losses due to evaporation.
- *Clothing* Clothes will act as insulators; the thicker they are, the more the fall in temperature will be delayed. This is also true for any kind of covering found on or around the body (e.g. subjects dying in bed and found lying under sheets and blankets).
- *Immersion in water* A corpse immersed in still water will lose heat mainly by conduction, at a rate several times higher than in air (because water is a much better heat conductor); the temperature decay will be even more accelerated if the water is flowing.

Henssge's nomogram probably constitutes the most elaborate and easy-to-use system developed to establish the time of death from body cooling, taking into account the main influencing factors. This three-parameter abacus (body temperature, ambient temperature, body mass) furnishes a crude estimation of the time since death, corresponding to the 'average' situation of a naked corpse lying in still air. Depending on the case, this preliminary value may be adjusted using corrective factors from 0.35 (for a naked body immersed in flowing water) to 2.40 (for a clothed body coved by a thick bedspread). Another major feature of this nomogram is that it provides the estimation together with a confidence interval (corresponding to a permissible variation of 95%); for an adult weighing 70 kg, this interval is ±2.8 h during the first 15 hours post mortem.

Many authors have proposed alternative solutions to enhance the accuracy of this technique, e.g. repeated or continuous postmortem temperature measurements over several hours, or the measurement of central temperature by invasive techniques (intrahepatic or subhepatic temperature taken by abdominal stab, or intracerebral temperature measured by inserting the probe through the orbit). These methods are difficult to apply to the routine investigation of scenes of death; in addition, none of them has proved to be clearly better than the single measurement of rectal versus environmental temperature.

Rigidity

Cadaveric rigidity (rigor mortis) is the consequence of irreversible and complex physicochemical changes occurring in muscle proteins after death, including conversion of muscle glycogen to lactic acid. Based upon many observations, this phenomenon has been proposed as an estimator of the time since death for about 200 years (its first scientific description was that of Nysten in 1811).

Rigidity usually develops sequentially and follows a descending pattern, the so-called Nysten's law: it affects successively the muscles of the face (especially those of the eyelids and lower jaw) and of the neck, then the trunk and upper limbs, and finally the lower limbs. It generally begins 3–4 h after death, is fully established after 8–12 h, remains unchanged for up to 36 h post mortem, and then disappears in 2–3 days – in most cases when putrefaction becomes patent. Breaking the rigidity manually (bending a joint by force, against the resistance of rigidity) will make it reappear if death occurred less than 8–12 h earlier; if the rigidity is complete, it will not reappear once broken down by force.

There are, however, many exceptions to this rule as far as both the sequential development and its time course are concerned. Numerous factors have been shown to affect, to a greater or lesser extent, the rate of onset and duration of rigor, or its intensity:

- *Temperature* is probably the more decisive parameter, as heat accelerates rigor whereas cold slows it down.
- *Violent exercise prior to death* may hasten the onset as well as disappearance of rigidity.
- *Muscular development:* the more muscular the subject, the greater the postmortem rigidity. On the other hand rigor mortis may be very moderate in children, emaciated people or in the elderly.

- *Cause of death:* rigidity may develop early, or even instantaneously ('cadaveric spasm'), in fatalities involving convulsant drugs (strychnine), electrocution, traumatic lesions affecting the central nervous system, or when death is preceded by extreme psychological stress (homicides, violent suicides). On the contrary, it may be delayed in some asphyxial deaths (carbon monoxide, hanging) or in deaths caused by massive hemorrhage.

In addition, the stage of rigor is generally evaluated subjectively, using criteria that vary from one author to another. Special devices based upon dynamometers have been proposed to measure rigidity with greater accuracy but they proved inconvenient and are never employed in casework.

Lividity

Postmortem lividity (hypostasis, livor mortis) is a plurifocal staining of the skin, usually in the form of a more or less intense purple discoloration, due to the gravitational settling of blood in vessels after the circulation has ceased. It always develops at the lowest parts of the cadaver, which depends upon its posture after death: in a body lying in the usual supine position, lividity predominantly affects the nape of the neck and the posterior aspects of trunk and limbs. By contrast, it does not appear on skin areas exposed to pressure, e.g. areas in contact with the underlying supporting surface (shoulder blades, buttocks, calves and heels for a body in the supine position), or areas squeezed by tightly fitting clothing (belts, underwear elastic).

Lividity often becomes perceptible within 3–4 h of death and progressively develops in surface area and colour intensity to attain its maximum degree 8–12 h postmortem. During this early period it is still mobile, i.e. it may be locally displaced by thumb pressure on the skin, or it can move (partly or completely) to other regions of the corpse if the body position is altered. After 12–15 h, postmortem hypostasis becomes 'fixed' and thus can no longer be displaced by external action. It will then remain unchanged until masked by the generally darker discoloration resulting from putrefaction.

Although this sign is present in almost all bodies (with the exception of those dying of massive hemorrhage), the time course of lividity exhibits considerable intersubject variability – probably to a greater extent than most other estimators of time of death in the early postmortem period. In addition, its quantitative measurement is imprecise (some authors use colorimetric tables) and in most routine cases its estimation remains largely subjective.

Potassium Concentrations in Vitreous Humor

Vitreous humor – the transparent, gelatinous substance that fills the posterior chamber of the eye – is an interesting medium for postmortem biochemical studies owing to its anatomical location and its relative resistance to bacterial contaminations over the first week after death.

The relationship between vitreous potassium levels and the time since death has been well studied for about 35 years. During life, potassium concentration is low in the vitreous humor but much higher in the peripheral tissues of the eye (vitreous layer, retina). This electrolytic imbalance results from energy-consuming, vital cell activities (active membrane transport, selective membrane permeability), the postmortem cessation of which leads to progressive reversal of the potassium gradient, with the consequence of a rise in vitreous concentration. Based on experiments on large series of cadavers, many authors have established linear relationships between vitreous potassium and time since death. Of the different equations proposed, the most popular are:

- Sturner's formula: $TSD = 7.14\,[K^+] - 39.1$
- Madea's formula: $TSD = 5.26\,[K^+] - 30.9$

where TSD is the time since death (hours) and $[K^+]$ is the potassium concentration ($mmol\,l^{-1}$).

Sampling

Vitreous humor is easy to collect but some rules have to be followed to avoid perturbances of the biochemical results. The fluid should be taken by puncture of both eyes at the external angle, using a syringe with intravenous or intramuscular needle. Suction *must* be gentle (do not use Vacutainer-type vials) to prevent aspiration of fragments from the retina or choroid; the samples should be centrifuged to eliminate cell debris. The time of collection must be carefully noted. Only colorless, crystal-clear vitreous is suitable for potassium analysis. Presence of blood, cell remains, or a brown-greenish color due to putrefaction dictates the sample to be withdrawn. If the potassium determination cannot be performed immediately, the samples should be stored frozen at $-18°C$.

Advantages and limitations

The main advantage of the vitreous potassium method is that it may be carried out up to 5–7 days after death, whereas most other estimators (especially body cooling) are useful only within the first 24 h. Unfortunately it also has some limitations:

- A major difficulty is presented by the 95% confidence interval of the technique which is always large and, in addition, somewhat different from author to author: for the first 24 h postmortem it ranges from ± 4 to ± 12 h, and for the period up to 100 h postmortem from ± 9.5 to ± 40 h.
- The rise of vitreous potassium concentrations may be strongly affected by the ambient temperature as well as by endogenous factors (age at death, duration of agony, etc.).
- At a given time, potassium concentrations may differ significantly between each eye (deviations can exceed 10% of the mean value of both eyes).
- All death time estimations by the vitreous potassium method assume that the antemortem period was not associated with electrolytic perturbations – which may be precisely the case in many deaths preceded by an agony of significant duration. Some authors have recommended the use of vitreous urea and/or creatinine in order to identify and reject subjects with possible antemortem electrolyte imbalance. By eliminating subjects with vitreous urea > 70 mg dl^{-1} or vitreous creatinine > 1 mg dl^{-1}, it has been stated that the 95% confidence interval for the period up to 120 h post mortem might be reduced to ± 15 h. If confirmed, this would represent the most accurate estimation of the time since death that could be obtained by any existing method.

Other Biochemical Markers

Since the 1950s, a number of other biochemical markers have been investigated to assess their potential as estimators of time of death:

- *Blood markers:* electrolytes (sodium, potassium, calcium, magnesium, phosphorus, chloride), glucose, lactic acid, urea, creatinine, cholesterol, triglycerides, apolipoproteins, hormones (cortisol, adrenalin, thyroxin, thyroid-stimulating hormone, insulin), enzymes (phosphatases, amylases, phosphoglucomutase), blood gases and pH, cleavage of the C3 component of complement.
- *Vitreous markers:* electrolytes (sodium, calcium, magnesium, chloride), urea, pH.
- *Cerebrospinal fluid markers:* electrolytes (potassium, magnesium, sodium, calcium, phosphorus, chloride).
- *Pericardial fluid markers:* electrolytes, cholesterol, glucose, lactic acid, enzymes (transaminases).
- *Muscle markers:* creatinine, enzymes (creatine phosphokinase).
- *Lung markers:* surfactant phospholipids.

Although experimental studies have shown that, in many cases, positive or negative correlations exist between the postmortem evolution of these markers and the time elapsed since death, their relative coarseness make them unsuitable for practical casework.

Supravital Reactions

Following cessation of the circulation, ischemia in organs and tissues leads to reversible, then irreversible changes affecting their structure and function. The time course of these phenomena is, however, very different, depending on the tissues; for example, brain cortex structures undergo definitive alterations after a few minutes, whereas other tissues (kidney, skeletal muscle) may tolerate prolonged ischemia for up to several hours. This intermediate period, beginning at brain death and lasting until cell activity has definitely ceased in the whole organism, is sometimes called the 'supravital period' because external stimuli applied to the corpse may induce life-mimicking, observable reactions. The supravital reactions that may be of interest in estimating the time since death are:

1. *Mechanical excitability of the skeletal muscle*, including:

   a Tendon reaction (Zsako's phenomenon), a contraction of the whole muscle due to propagated excitation following a mechanical stimulation. This can be obtained, for instance, by striking the lower third of the thigh 4–5 fingerbreadths above the patella with a reflex hammer, resulting in an upward movement of the patella due to contraction of the whole quadriceps muscle. Zsako's phenomenon is particularly transient, as it usually cannot be observed beyond 2–3 h postmortem.

   b Idiomuscular contraction, a localized muscular contraction (bulge) at the point of stimulation, demonstrated, for instance, by striking the biceps muscle of the arm with a reflex hammer. This can be observed several hours after cessation of Zsako's phenomenon.
2. *Electrical excitability of the skeletal muscle*. This is generally investigated on the muscles of the face. Needle electrodes are inserted through the skin (at the nasal part of the upper eyelid, or on both sides of the mouth) and electrical impulses, provided by a portable generator, are applied. The reaction – a contraction of one or several muscles – may be semiquantitatively quoted according to the strength of the contraction and its extension to areas distant from the electrodes, both of them diminishing as time since death increases.
3. *Pharmacological excitability of the iris muscle*. Modifications of the pupil diameter can be observed following the administration of miotic

(acetylcholine) or mydriatic (norepinephrine, atropine) solutions, which may be instilled on to the cornea or injected subconjunctivally at the limbus of the cornea.

Supravital reactivity may be of value for establishing the time of death, but only during the first hours of the postmortem period (none of the above-mentioned phenomena proved observable beyond 12–15 h postmortem). Since the muscular responses obtained can be estimated only semiquantitatively, their investigation requires very experienced operators. The equipment needed for electrical stimulation also limits the applicability of this technique for routine scene-of-death investigation.

Putrefaction

Putrefaction is the destruction of the soft tissues of the cadaver by the action of bacteria and enzymes. Its evolution can be divided into five successive periods:

- *Initial decay* (up to 36–72 h post mortem): the corpse still appears fresh externally but internally it begins to decompose, owing to the combination of enzymatic autolysis and bacterial proliferation from the intestine.
- *Early putrefaction or green putrefaction* (up to 1 week post mortem): this period often begins with a greenish discoloration of the right iliac fossa, which subsequently spreads to the whole anterior abdominal wall and the rest of the trunk, then to the neck, face and finally to the limbs (upper and lower limb extremities typically putrefy last). Other signs include skin blisters; abdominal/scrotal swelling with gas; oozing of putrefactive fluids by mouth and nostrils (not to be mistaken for blood); and the typical odor of decaying flesh.
- *Black putrefaction* (up to 1 month post mortem): the cadaver exhibits a flesh of creamy consistency with exposed parts turning black, especially at the head and face. Skin decomposition results in generalized epidermal detachment, and the nails fall off. The abdomen collapses as gases escape. The odor of decay is at its maximum.
- *Butyric fermentation* (up to 2 months post mortem): this stage is marked by a progressive drying of the cadaver and the occurrence and proliferation of mold. The odor is less offensive and becomes typically cheesy.
- *Dry decay*, then *skeletonization* (months to years): final drying of the cadaver and progressive disappearance of the remaining soft tissues. Time-related changes affecting the remains at this stage are slow compared with those of the preceding periods.

Unfortunately, this chronology of putrefactive events is only indicative of what may occur, as enormous and generally unpredictable intersubject variability exists. Ambient temperature obviously influences the rate of putrefaction to a major extent, but many other factors may intervene, such as body corpulence (obese subjects putrefy more quickly than slender ones), antemortem diseases or circumstances of death (sepsis or edemas during the period preceding death may hasten decomposition). Furthermore, there are some variants of the ‘classical’ succession of decomposition stages, depending on environmental circumstances: adipocere formation (in warm, damp, preferably anaerobic conditions) and mummification (in hot and dry environments, especially if air is moving). For these reasons, and however great the experience of the forensic investigator, the putrefactive phenomena affecting a cadaver cannot be considered as reliable markers of the time elapsed since death.

Entomology

After death, the tissues of animals become attractive to a large variety of insects and other invertebrates. These may be classified into four groups:

- *Necrophagous species*: invertebrates that feed on the corpse itself.
- *Predators and parasites of the necrophagous species*: these species do not feed directly on the corpse.
- *Omnivorous species*: invertebrates that feed both on the corpse and on the other arthropods present.
- *Adventive or opportunistic species*: invertebrates present on the corpse by chance, or using it as an extension of their usual environment (e.g. as a shelter, nest) without feeding on it.

The use of entomologic markers for determining the time of death is based upon the long-established observation (first scientifically reported by Mégnin in 1894) that insects and other arthropods feeding on a corpse follow a specific faunal succession associated with the various stages of decay. Estimations of time of death require an accurate recognition of the sometimes numerous species present on a corpse, or its surroundings, in their different immature (eggs, larvae, puparia) or adult stages of growth, together with an extensive knowledge of their specific rates of development according to environmental parameters (season, temperature, humidity, etc.). Such investigations are difficult and can be undertaken only by experienced, full-time specialists in forensic entomology. Provided this is the case, this technique currently constitutes the only approach for estimating the time of death of a putrefied cadaver with some accuracy –

in some cases, to within a few days, even in deaths obviously dating back for months.

Conclusions: Combined Methods

As shown by this brief survey, all methods proposed for the determination of the time since death remain relatively inaccurate, even when applied to the very early postmortem period. This limitation is a consequence of the huge interindividual variability affecting all postmortem changes, which is in large part attributable to the existence of many complicating factors, endogenous as well as exogenous: location of the body (exposed to atmosphere/buried/immersed); environmental temperature and other seasonal/climatic conditions (wind, humidity, exposure to sun); nature of soil; body corpulence; posture of the corpse; presence/nature of clothes or coverings; antemortem diseases; duration of the process of dying; causes and circumstances of death. Another reason for this poor level of accuracy is that many of the criteria used for the estimation are subjectively evaluated instead of being objectively measured. Finally, reporting to the police the time at which death is deemed to have occurred is of little interest, scientifically questionable and potentially dangerous if the crude estimation (usually a mean value obtained by computation of one or several tables or formulas) is not furnished together with an interval of confidence.

As emphasized by several authors, partial remedies may consist in:

- understanding and taking into account the major complicating factors acting on each criterion;
- using objective measurements instead of subjective evaluations;
- rejecting estimators that do not allow the calculation of confidence limits for the time of death (e.g. lividity, putrefaction), or using them only as adjustment parameters for more reliable methods.

The combination/integration of different estimators has been proposed to narrow down the margins of error associated with single methods. For practical casework, Henssge and coauthors use special charts in which they arrange the following parameters:

- body cooling (interpreted using Henssge's nomogram with corrective factors);
- external examination of lividity and rigidity;
- electrical and mechanical excitability of facial muscles;
- chemical excitability of the iris.

Despite these recent improvements and unless some revolutionary technique – at present unforseeable – is introduced in the future, the great accuracy sometimes depicted in movies and detective novels ('Your Honour, death occurred yesterday at 5.47 p.m.') will remain an unrealistic mirage. The determination of the time since death is a difficult and complex task, frequently disappointing in the modest results it produces. Obviously, this work requires both scientific competence and field experience – but it also demands a large dose of humility.

See also: **Causes of Death**: Postmortem Changes. **Pathology**: Postmortem Changes; Postmortem Interval.

Further Reading

Coe JI (1974) Postmortem chemistry: practical consideration and a review of the literature. *Journal of Forensic Sciences* **19**:13–42.

Gordon I, Shapiro HA and Berson SD (1988) *Forensic Medicine: A Guide to Principles*. Edinburgh: Churchill Livingstone.

Hansson LMA, Uotila U, Lindfors R and Laiho K (1966) Potassium content of the vitreous body as an aid in determining the time of death. *Journal of Forensic Sciences* **11**:390–394.

Henssge C (1988) Death time estimation in case work – I. The rectal temperature time of death nomogram. *Forensic Science International* **38**:209–236.

Henssge C, Madea B and Gallenkemper E (1988) Death time estimation in case work – II. Integration of different methods. *Forensic Science International* **39**:77–87.

Henssge C, Knight B, Krompecher T, Madea B and Nokes L (1995) *The Estimation of the Time Since Death in the Early Postmortem Period*. London: Edward Arnold.

Knight B (1991) *Forensic Pathology*. London: Edward Arnold.

Madea B and Henssge C (1990) Electrical excitability of skeletal muscle postmortem in casework. *Forensic Science International* **47**:207–227.

Randall B (1997) *Death Investigation: The Basics*. Chicago: Galen Press.

Smith KGV (1986) *A Manual of Forensic Entomology*. Ithaca: Cornell University Press/Comstock Publishing Associates.

Sturner WQ and Gantner GE (1964) The postmortem interval, a study of potassium in the vitreous humour. *American Journal of Clinical Pathology* **42**:137–144.

Toolmarks *see* **Pattern Evidence**: Tools.

TOXICOLOGY

Contents
Overview
Equine Drug Testing
Inhalants
Interpretation of Results
Methods of Analysis – Ante Mortem
Methods of Analysis – Post Mortem

Overview

O H Drummer, Victorian Institute of Forensic Medicine and Monash University, Melbourne, Australia

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0543

Introduction

Toxicology is the science of poisons; when applied to medicolegal proceedings, the terms forensic toxicology or analytical toxicology are often used. A forensic toxicologist is concerned with the detection of drugs or poisons in samples and is capable of defending the result in a court of law. This distinction from an ordinary analytical toxicologist is important, as a conventional toxicologist is mainly concerned with the detection of substances, and may not understand the specific medicolegal requirements in forensic cases.

The process of conducting toxicology is similar to other analytical disciplines, in that sufficiently suitable analytical techniques need to be employed, which are appropriately validated. The conduct of suitable quality assurance is important to assure the analyst and clients of the quality of the result. These issues are discussed in this overview, while in other articles specific issues of techniques, specimens and interpretation are further discussed.

Applications of Forensic Toxicology

Forensic toxicology has a number of applications. Traditionally, it is used in death investigations. It provides clinicians with information of a possible drug taken in overdose, or authorities investigating a sudden death or poisoning with information on the possible substances(s) used. Ultimately toxicology testing results will assist the medical practitioner, coroner or medical examiner in establishing the evidence of drug use, or by refuting the use of relevant drugs.

Toxicology testing is also important in victims of crime, or in persons apprehended for a crime. Drugs may have been given by the assailant to reduce consciousness of the victim, such as in rape cases. These drugs include the benzodiazepines (e.g. Rohypnol) and γ-hydroxybutyrate (GHB). Toxicology also establishes if any drug was used by the victim and which may have affected consciousness or behavior. Defendants arrested shortly after allegedly committing a violent crime may be under the influence of drugs. It is vital, therefore, that toxicology testing is conducted (on relevant specimens) to establish the extent of drug use, as allegations of drug use and its effect on intent or clinical state may be raised in legal proceedings.

Forensic toxicology is also used in employment drug testing and in human performance testing. The former category relates to the detection of drugs of abuse in persons in a place of employment, prior to being hired by an employer, or even a person in detention, such as in a prison. Human performance testing relates to the detection of drugs that might have increased (usually) performance in athletic events. This may even apply to animals such as horses. Specimens used in these cases are usually urine, although hair is increasingly used to provide a longer window of opportunity.

Initial Tests and Confirmation

The foremost goal in forensic toxicology is the need to provide a substantial proof of the presence of a substance(s). The use of conventional gas chroma-

tography (GC), thin-layer chromatography (TLC) or high-performance liquid chromatography (HPLC) would not normally be sufficient to accept unequivocal proof of the presence of a chemical substance. Two or more independent tests are normally required, or the use of a more powerful analytical test, such as mass spectrometry (MS) is often preferred. Because of the need to perform a rigorous analysis, the analytical schema is often broken up into two steps. The identification stage is termed the screening or initial test, while the second analytical test is the confirmation process. The confirmation process often also provides a quantitative measure of how much substance was present in the sample, otherwise a separate test is required to quantify the amount of substance present in the specimen (see later). In all processes it is important that no analytical inconsistency appears, or a result may be invalidated (**Fig. 1**). For example, in the identification of codeine in a blood specimen, an immunoassay positive to opiates is expected to be positive when codeine is confirmed. The apparent detection of a drug in one analytical assay, but not in another, means that the drug was not confirmed, providing both assays are capable of detecting this drug.

Table 1 lists common techniques used in screening and confirmation assays. While MS is the preferred technique for confirmation of drugs and poisons, some substances display poor mass spectral definition. Compounds with base ions at mass:charge ratios of less than 100, or with common ions such as m/z 105 and with little or no ions in the higher mass range, are not recommended for confirmation by MS alone. Derivitization of a functional group to produce improved mass spectral properties can often be successful. Common derivatives include perfluoroacyl esters, trimethyl silyl ethers, etc. Alternatively,

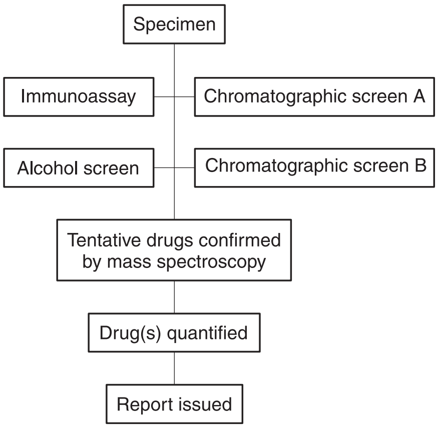


Figure 1 Identification, confirmation and quantification processes in forensic toxicology.

Table 1 Screening and confirmation techniques

Screening tests	*Confirmation tests*
Immunoassays	MS (LC, GC, CE)
Spectroscopy (UV, F, etc.)	Second chromatographic test
HPLC (UV, F, ECD, CD)	HPLC (DAD)
GC (FID, NPD, TD)	AAS
CE (UV, F)	ICP-MS
AAS, colorimetric tests	

AAS, atomic absorption spectroscopy; CD, conductivity detection; CE, capillary electrophoresis; DAD, photodiode array detector; ECD, electrochemical detector; F, fluorescence; FID, flame ionization detector; GC, gas chromatography; HPLC, high-performance liquid chromatography; ICP-MS, inductively coupled plasma MS; LC, liquid chromatography; MS, mass spectrometry; NPD, nitrogen–phosphorus detector; TD, thermionic detector; UV, ultraviolet.

reliance on other chromatographic procedures can provide adequate confirmation. It is important when using any chromatographic procedure (HPLC, GC, capillary electrophoresis (CE), etc.), that the retention time of the substance being identified matches with that of an authentic standard.

Some apparent analytical inconsistencies may provide important forensic information. For example, if a result for opiates is negative in urine, but positive in blood, it is possible that heroin (which is rapidly metabolized to morphine) was administered shortly before death, and therefore metabolites had not yet been excreted. This situation is often found in heroin users suffering an acute sudden death, in whom substantial urinary excretion has not yet occurred.

Common Drugs and Poisons

The most common drugs and poisons are clearly the initial targets of any forensic toxicological analysis, particularly if no specific information is available to direct the investigation. The most common substances can be categorized as fitting into four classes: alcohol (ethanol), illicit drugs, licit (ethical) drugs, and the nondrug poisons. An example of the distribution of drugs in various types of coroner's cases is shown in **Table 2**. These data are likely to be similar throughout developed countries.

Alcohol is the most frequent finding in many countries, and, when detected, can play an important role in any investigation because of its ability to depress the central nervous system (CNS). At best, alcohol will modify behavior, causing disinhibition and possible aggression; at worst, it can cause death, either by itself, or in combination with another drug.

Illicit drugs include the amphetamines, barbiturates, cocaine, heroin and other opiates, cannabis,

Table 2 Incidence of drugs in various types of death (%)[a]

Type of death	*Ethanol*	*Opioids*[b]	*Benzodiazepines*	*Stimulants*[c]	*Cannabis*	*Antipsychotics*
Natural death	15	13	9.4	1.4	2.3	2.6
Homicides	38	11	11	4.0	16	0
Drivers of motor vehicles	27	6.2	4.3	4.3	16	<1
Nondrug-related suicides	33	10	21	2.9	13	2.1
Licit drug deaths	40	41	59	3.2	8.0	13
Illicit drug deaths	35	96	61	7.1	38	5.4
All cases	27	20	20	3.1	12	3.2

[a] Taken from 2000 cases.
[b] Includes codeine and propoxyphene.
[c] Includes legal stimulants, amphetamines and cocaine.
Data produced with permission from 1997 files of the Victorian Institute of Forensic Medicine.

phencyclidine, designer fentanyls and lysergic acid diethylamide (LSD). It should be borne in mind that some illicit drugs also have medical uses in some countries. Cocaine is used in some forms of facial and nasal surgery, amphetamine is used to treat narcolepsy and attention deficit syndrome, and cannabis is used (among other indications) to reduce nausea following chemotherapy.

Ethical drugs include the whole range of prescription and over-the-counter drugs used in the treatment of minor to major ailments. Those of most interest include the antidepressants, major tranquilizers, narcotics and other forms of pain relievers, anticonvulsants, etc. Since these drugs are widely prescribed, this is by far the most common drug category encountered in toxicology. Each country will have its own list of registered drugs, hence laboratories will need to consider these as a matter of priority over other members of a particular class available elsewhere. For example, most countries only have a relatively small number of benzodiazepines registered for medical use, whereas over 35 are available throughout the world. From time to time laboratories will be required to consider drugs not legally available in their own countries because of illicit supplies arriving in, or tourists visiting, their country.

The nondrug poisons include, most commonly, organophosphates and other pesticides, carbon monoxide, hydrogen cyanide and cyanide salts, and volatile substances (petrol, gas, kerosene, etc.). (Carbon monoxide and hydrogen cyanide are gases emitted by fires and are therefore frequently detected in victims of fire.) Other poisons include heavy metals (arsenic, mercury, thallium, etc.), plant-derived poisons (hyoscine from belladonna, coniine from hemlock, etc.), strychnine, and toxins such as venoms. Performance-enhancing drugs such as the anabolic steroids may also be considered in some instances. Clearly this list is potentially unending, although some chemicals are more readily available to certain occupational groups than others and are dependent on national regulations within countries. In a review of 10 years of forensic cases from Victoria, Australia (Victorian Institute of Forensic Medicine toxicology laboratory 1989–1998), the distribution of unusual poisons was as shown in **Table 3**. Clearly, the distribution of unusual drugs and poisons will vary from country to country.

Scope of Testing Protocols

As the previous sections indicate, cases may involve a variety of ethical and illicit drugs, or unusual poisons. Worldwide experience also shows that forensic cases often involve more than one drug substance. A survey of drug-related deaths shows three or more drugs are present in more than 70% of cases. High rates of multiple drug use are also found in perpetrators and victims of violent crimes and suicides, and often also in accidents and road crashes.

Table 3 Incidence of poisons in coroners' cases

Poison	*10 year incidence*[a]
Organophosphates	19
Butane and other hydrocarbons	12
Other pesticides/herbicides	12
Solvents (methanol, chloroform, etc.)	9
Strychnine	7
Potassium cyanide	6
Plant-derived poisons	5
Ethylene glycol	3
Heavy metals	2
Potassium	2
Others	11

[a] Taken from over 20 000 coroners' cases 1989–1998.

It is also well known by forensic toxicologists that the information provided to the laboratory concerning possible drug used may not accord with what is actually detected. It is therefore strongly recommended that laboratories provide a systematic approach to their toxicology cases and include as wide a range of common ethical and illicit drugs as feasible. This approach is termed systematic toxicology analysis (STA). A laboratory using this approach would normally include a range of screening methods, often incorporating both chromatographic and immunological techniques. Drug classes such as alcohol, analgesics, opioid and nonopioid narcotics, amphetamines, antidepressants, benzodiazepines, barbiturates, cannabis, cocaine, major tranquilizers (antipsychotic drugs) and other CNS depressant drugs would be included.

The incorporation of a reasonably complete range of drugs in any testing protocol is important because many of these drugs are mood-altering, and can therefore affect behavior as well as affecting the health status of an individual. Persons using benzodiazepines, for example, will be further affected by cocaine, amphetamines, and use of other CNS depressant drugs. The toxic concentrations of drugs are also influenced by the presence of other potentially toxic drugs. For example, the fatal dose for heroin is affected by the concomitant use of alcohol and other CNS depressant drugs.

Specimens

It is essential that the relevant specimens are taken whenever possible, as re-collection is rarely an option. The preferred specimens collected in forensic toxicology will, of course, depend on the nature of the case. In general, a blood specimen is a minimum requirement, although specimens such as urine can be useful for screening and to check for the use of drugs two or more days prior to sampling.

Hair can provide an even longer memory of drug intake, lasting up to several months, depending on the length of hair. Drugs are usually incorporated into the growing root and appear as a band in the hair shaft when it externalizes from the skin. This process can therefore provide a history of when exposure to a drug or poison has occurred. Most drugs and poisons are incorporated into hair, although the extent will depend on the physiochemical properties of the substance. Basic drugs are often found in higher concentrations than acidic drugs, and invariably the parent drug is present rather than metabolites. For example, cocaine and the heroin metabolite 6-monoacetylmorphine are more likely to be found in hair of cocaine and heroin users than their corresponding metabolites found in blood and urine (benzoylecgonine and morphine). Unfortunately, some drug will be absorbed into the hair from skin secretions adjacent to the hair follicles, and may even be incorporated from external contamination. Care and treatment of hair, hair color, such as washing, use of dyes and bleaches, etc., will also affect the concentration of drug in hair. Consequently, any interpretation of drug content in hair needs to take these factors into account. The advantages of other specimens are described elsewhere.

Courts and other legal processes usually require proof that the laboratory has taken all reasonable precautions against unwanted tampering or alteration of the evidence. This applies to specimens and to physical exhibits used by the laboratory in their toxicology investigations. (The term ‘exhibit’ applies both to specimens and physical items, such as tablets, syringes, etc.) Consequently, it is essential that the correct identifying details are recorded on the exhibit or specimen container, and an adequate record is kept of persons in possession of the exhibit(s). When couriers are used to transport exhibits, the exhibit must be adequately sealed to prevent unauthorized tampering.

Procedures are available to assist laboratories in establishing suitable chain-of-custody records.

General Techniques

The techniques available for the detection of drugs in specimens collected post mortem are essentially identical to those for specimens collected ante mortem. These range from commercial kit-based immunoassays (enzyme multiplied immunoassay technique (EMIT), fluorescence polarization immunoassay (FPIA), cloned enzyme donor immunoassay (CEDIA), radioimmunoassay (RIA), etc.), traditional TLC, to instrumental separation techniques, such as HPLC, GC and CE. MS is the definitive technique used to establish proof of structure of an unknown substance, and can be linked to GC, HPLC and more recently to CE.

The use of appropriate extraction techniques is critical to all analytical methods. Three main types of extractions are used: liquid–liquid, solid-phase and direct injection. Traditionally, liquid techniques, in which a blood or urine specimen is treated with a buffer of an appropriate pH followed by a solvent capable of partitioning the drug out of the matrix, have been favored. Solvents used include chloroform, diethyl ether, ethyl acetate, toluene, hexane, various alcohols and butyl chloride, and mixtures thereof.

The solvent is then isolated from the mixture and either cleaned up by another extraction process or evaporated to dryness.

Solid-phase techniques are becoming increasingly favored, as mixed phases offer the ability to extract substances of widely deferring polarity more readily than with liquid techniques. Often less solvent is used, or simple hydroalcoholic systems can be employed, rather than potentially volatile or inflammable solvents.

Direct injection techniques into either GC or HPLC instruments bypass the extraction step, and can offer a very rapid analytical process. In GC, solid-phase microextraction (SPE) is most commonly used, while HPLC tends to require the use of precolumns, which are back-flushed with the use of column switching valves.

Quality Assurance and Validation

An essential part of any form of toxicological testing is validation and quality assurance. It is important that the method used is appropriately validated; that is, it has been shown to accurately and precisely identify the substance(s) detectable, there is little or no interference (from other drugs or from the matrix) with the specimens used, and that a useful detection limit has been established. Moreover, it is essential that the method is rugged and will allow any suitably trained analyst to conduct the procedure and achieve the same results as another analyst. To achieve these aims, it will be necessary to test the method in the laboratory over several assays, with varying specimen quality, before claiming that a full validation has been conducted.

It is recommended that internal quality controls with each batch of samples be included to enable an internal check of the reliability of each assay. These controls contain known drugs, at known concentrations. Suitable acceptance criteria are required for these controls before results of unknown cases can be accepted and released to a client. Acceptance criteria vary, depending on the analyte and application. For example, blood alcohol estimations have acceptance criteria less than 5%, while postmortem blood procedures may be 10–20%. (Normally the coefficient of variation of the mean (CV) is calculated as a standard deviation divided by the mean of the result.)

An important feature of analytical assays in forensic toxicology is the use of internal standards. These are drugs of similar chemical and physical characteristics as the drug(s) being analyzed, and, when added at the start of the extraction procedure, provide an ability to negate the effects of variable or low recoveries from the matrix. Hence, even when recoveries are low, the ratios of analyte and drug are essentially the same as for situations of higher recovery. An ideal recovery marker is when the internal standard is a deuterated analog of the analyte. When deuterated internal standards are used, it may not be necessary to match the calibration standards with the same matrix as the unknown samples. It is important, however, that absolute recoveries are reasonable, i.e. at least over 30%. This ensures less variability between samples and optimizes the detection limit.

From time to time it will be important to run unknown samples prepared by another laboratory, or a person not directly involved in laboratory work, to establish proficiency. These are known as proficiency programs or quality assurance programs. These trials are often conducted with many other laboratories conducting similar work, and provide an independent assessment of the proficiency of the laboratory to detect (and quantify) specific drugs. The performance of the laboratory should be regularly assessed from these results, and any corrective action implemented, if appropriate. This process provides a measure of continuous improvement, an essential characteristic of any laboratory. There are a number of collaborative programs available throughout the world. The College of American Pathologists (CAP) organizes an excellent series of proficiency programs in forensic toxicology.

The international (TIAFT) and American (SOFT) societies of forensic toxicology provide guidelines on the conduct of analytical assays and quality assurance of assays.

Postmortem Artifacts in Analysis

The event of death imparts a number of special processes that affect the collection and analysis of specimens obtained at autopsy. These include postmortem redistribution, in which the concentration of a drug in blood has been affected by diffusion of the drug from neighboring tissue sites and organs, such as stomach contents. This is minimized, but not arrested, using peripheral blood from the femoral region. Even liver concentrations are affected by diffusion from intestinal contents or from incomplete circulation and distribution within the liver. Some drugs are metabolized after death – nitrazepam, flunitrazepam, heroin, aspirin, etc. Substances such as ethanol and cyanide may even be produced by bacterial processes in decomposing bodies.

Estimation of Dose

A common request from legal counsels and police is to estimate a dose from a blood or tissue concentration. This may relate to determining likely intent from an ingestion (or injection), or simply to rationalize the circumstances to specific amounts of drugs used.

Dose can be estimated from knowledge of the volume of distribution (V_d) of drug. The calculation multiplies the blood concentration by the V_d corrected for the body weight of the person. Unfortunately, this calculation assumes one V_d for all persons, and, importantly, assumes that equilibrium has been established at the time of drug ingestion. This is rarely the case in toxicology cases, as recent drug ingestion is common. The calculation also fails to account for unabsorbed drug (and excreted drug) and may be severely affected by postmortem processes.

The variation in blood concentration at a specified time from a standard dose of drug is well known in clinical pharmacology, even in controlled situations. Therefore, estimating dose is not recommended unless these factors are considered and a range of doses is computed. Occasionally, it may be possible to compare blood (and tissue) concentrations to other cases in which doses were known, or by measuring the body burden in several tissues, including muscle and fat. Analysis of gastric and intestinal drug content will assist in this process, and also provide information on the route and time of ingestion.

Interpretation of Toxicological Results

Interpretation of any toxicological result is complex. Consideration must be given to the circumstances of the case, and in particular what significance may be drawn from the toxicology. For example, the finding of a drug in potentially toxic concentrations in a person killed by a gunshot wound to the head cannot reasonably lead to the conclusion that drugs caused the death. On the other hand, the absence of an obvious anatomical cause of death will lead investigators to consider the role of any drug use. Considerations must include the chronicity of drug use, the age of the person, the health of the person (presence of heart, liver, kidney disease, etc.), the use of multiple substances, and even genetic factors that may lead to a reduced metabolism.

Problems in Court Testimony

Forensic toxicologists and other analysts called to give evidence in court should consider that much of their technical evidence is beyond the ready comprehension of lay people in juries, legal counsel and judges. Restricting their evidence to understandable language and simple concepts is highly recommended.

A further problem relates to an assumption often made by legal counsel (and indeed other parties) that a toxicological investigation was exhaustive and all drugs and poisons were excluded in the testing processes. Most toxicology performed is restricted to a few analytical tests for a range of 'common drugs and poisons', unless the client (e.g. pathologist or police) has made a request to examine for (additional) specific chemicals. Analysts should make courts aware of the actual testing conducted and provide a list of substances incorporated in the investigation. Importantly, advice on any limitations applied to the interpretation of the analytical results should be supplied, e.g. poor quality specimens, postmortem artifacts, etc. Above all, toxicologists must restrict their evidence to those areas in which they claim expertise. Stretching their expertise to assist the court can lead to incorrect or misleading evidence, and damage the reputation of the expert.

See also: **Toxicology**: Methods of Analysis – Ante Mortem; Methods of Analysis – Post Mortem.

Further Reading

Baselt RH and Cravey RH (1996) In: *Disposition of Toxic Drugs and Chemicals in Man*, 4th edn. New York: Year Book.

de Zeeuw RA (1997) Drug screening in biological fluids. The need for a systematic approach. *Journal of Chromatography B Biomedical Sciences and Applications* 689:71–79.

de la Torre R, Segura J, de Zeeuw R and Williams J (1997) Recommendations for the reliable detection of illicit drugs in urine in the European Union, with special attention to the workplace. *Annals of Clinical Biochemistry* 34:339–344.

Drummer OH (1998) Adverse drug reactions. In: Selby H (ed.) *The Inquest Handbook*. Leichhardt, NSW: Federation Press.

Drummer OH (1999) Chromatographic screening techniques in systematic toxicological analysis. *Journal of Chromatography* 733:27–45

Karch S (ed.) (1998) In: *Drug Abuse Handbook*, Boca Raton, FL: CRC Press.

Moffatt AC (ed.) (1986) *Clarke's Isolation and Identification of Drugs*. London: Pharmaceutical Press.

United Nations (1997) *Recommended Methods for the Detection and Assay of Barbiturates and Benzodiazepines in Biological Specimens* ST/NAR/28. New York: United Nations.

Equine Drug Testing

L Rivier, C Cardis and **P Mangin**, University Institute of Legal Medicine, Lausanne, Switzerland

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0601

Introduction

The twentieth century saw the development of a multimillion dollar horseracing industry. For competitions, two main elements must be considered in the control of drugs in racehorses:

1. Legitimate treatment, supervised by a veterinarian, may be underway for illness or injury, so that the horse will be fit to perform.
2. Drugs may have been administered to the horse with the intent of altering its performance. This is tested for by analysing official test samples (urine and/or blood), which are taken before and/or after competition. In some countries, between-competition testing is also organized. Treated horses must not race until they have eliminated any traces of the administered substances and/or their metabolites.

This dual nature of therapeutic medication and drug abuse in racehorses created a difficult regulatory problem. The act of taking substances that influence the physical performance, and are usually banned, is now known as doping. Most racing jurisdictions have strict medication rules, similar to the control of drugs in human athletes, which prohibit the presence of any drug in an official test sample of urine, blood or other body fluids taken before or after a race. The most recent rules of the majority of governing bodies for horse sport conform roughly to those of the Fédération Equestre Internationale (FEI). Similar rules have been adopted by the International Federation of Horseracing Authorities (IFHA). These two bodies are the leading international equestrian sports associations. The regulations forbid the presence of any performance-enhancing substances in the body of the animal. The stated objective of such regulations is to protect fair competition or, more specifically, any person betting on horses of unknown or inconsistent performance owing to the influence of drugs. Racing analysts and veterinarians are the forensic and sport medicine experts who are responsible for monitoring the medication of competing horses. The prime responsibility of racing laboratories, together with racing authorities, is to protect the integrity of racing.

Context of Control

Owing to the proliferation of pharmaceuticals in the last 40 years, it has become difficult for the analyst to keep up with new drug developments. Even before new pharmacologically active substances are introduced on the market, whether for human or veterinary uses, they may appear in samples taken during doping controls. Comparing the work of the scientist who has to analyse urine and/or blood samples from competing horses with analysts dealing with clinical, toxicological or forensic samples shows obvious differences. In sport, the laboratory receives coded samples, with the only indication being the type of horse. No additional information is given; one is not provided with a case report, nor advised to look for any particular class of drug because of any evidence such as apparent intoxication. If a doping agent has been taken, the amount in blood is likely to be within the normal, that is the therapeutic, range of concentration. Obviously, we never encounter any drug overdose case where large concentrations of the drug and/or metabolites are typically present.

A limited volume of urine or blood is provided, generally 40–50 ml, and, for the large majority of cases, there will be no opportunity for a second, follow-up sample. Nevertheless, the exact identity of any banned substance present in the sample must be made unequivocally, and frequently within strict time limits of, usually, between 3 and 10 days, as specified in the contract agreed with the controlling authorities.

Since most of the samples are usually negative, i.e. are found not to contain any substance from a banned class, and since confirmatory tests are relatively complex, time-consuming and expensive, all samples are submitted to a range of screening tests. The screening tests usually rely on one or several techniques: thin-layer chromatography (TLC), gas chromatography (GC), high-performance liquid chromatography (HPLC), mass spectrometry, usually coupled with GC (GC-MS) or HPLC (LC-MS), and immunoassays (RIA and/or ELISA, for example). Only those samples that fail screening tests are then submitted to confirmatory testing. The screening tests are designed to be sensitive, and typically 20% of the samples require some form of confirmatory tests. The confirmatory test relies heavily on GC-MS or LC-MS, together with the use of chemical derivatization and/or chemical ionization to improve both GC and MS characteristics of the substance. The laboratory must frequently identify metabolites of the substance in urine, especially when the parent compound is undetectable. This is evidenced by comparison with with urine and/or blood samples obtained from controlled administration of banned substances in healthy horses.

How Horses are Tested

At each competition, judges or stewards can select for testing not only the winner of a race but also other horses, at random or depending on their individual performance during the race. There are no strict rules for the number of samples to be collected at each competition. The FEI states that, for each competition, a minimum of 5% of the horses shall be controlled. Most of the time, only the winner and an antepost favourite which has lost might be preferentially selected. An official test sample (urine or blood) is collected from the horse by the designated test inspector or under his or her supervision, in order to determine the presence of any prohibited substance. Ideally, this inspector shall be independent from the laboratory, and is responsible for the supervision of the horse before the withdrawal of the samples, the collection itself, the mailing of the samples and the establishment of the related documentation.

The term 'prohibited substance' means the substance, the metabolite(s) and the isomers or biological indicators of such substance (including any metabolite(s) originating externally), whether or not they are endogenous to the horse, and which are contained in the official list (see below). Horses may compete with the presence of certain substances in their tissues, body fluids or excreta, for which specific instructions to the laboratories have indicated threshold levels or ratios (as reported below), provided the concentration of the substance is not greater than that indicated. When the analysis of a sample gives a positive finding for a substance which may have been endogenously produced, the horse may be submitted to further examination. The same is true for the finding of a substance which could be related to food contamination. In view of the rapid development of new drugs and pharmacological agents, and of the changes in preparing horses for competition, the regulations governing the use of new products and new techniques are kept under continuous review and may be changed at any time.

Immediately after the race ends, the designated horse is conducted to a specially prepared retention area, where the animal is placed in a secure stall and watched until such time as the samples can be obtained with specially designed tools (sampling kits are usually approved by the controlling authorities). Urine is preferred as the laboratory has a better chance of detecting forbidden substances. It is also possible to get a larger volume and it is considered noninvasive. If the horse has not been able to provide any urine within 1 h of the end of the race, venous blood is taken by a veterinarian or any other qualified individual, in the presence of the test inspector. Once the official samples have been collected, the bottles are sealed and labeled with track name, date and a unique number, and kept in a refrigerator until dispatch to the laboratory. The bottles are mailed in a sealed box within 24 h, together with the corresponding documentation labeled in such a way that the accredited laboratory cannot identify the horses. The full documentation is kept apart in a safe place.

Penalties and Arbitration

After a positive test is reported, the racing commission notifies the trainer of the horse that has produced a positive test sample. A hearing is held to assess the penalty to be given, according to the commission's rules. The owner or trainer has the right to appeal the ruling to the racing commission. Any horse found to have a prohibited substance in any of its tissues, body fluids or excreta at an event is automatically disqualified, together with the competitor, from all competitions at the same event. The classification is adjusted accordingly. If the disqualified horse and competitor are members of a team, the rest of that team is not disqualified. Appeals may be directed to the Court for Arbitration in Sport (CAS) in Lausanne. Any emergency treatment to a competing horse just before an event must be certified, in writing, by the veterinarian, with details of the extraordinary circumstances. In cases of obvious illness or injury during an event, the jury will decide whether the horse may continue in that or subsequent competitions.

Owners or trainers may ask for the sample residue for independent testing to verify a positive test result. A sample residue is the urine or blood remaining after the analysis of an official sample has been completed. The independent analysis can be conducted at a referee laboratory, selected by the owner or trainer. Care should be taken selecting a referee laboratory, as the methods used by the referee laboratory might not be comparable to those used by the primary one. Further, drug breakdown or sample deterioration could cause different analytical results. Consequently, the analytical results of the referee laboratory may not be given equal weight as evidence at a hearing.

Establishing the Presence of Prohibited Substances

Objective

The objective of controlling the use of substances is based on the capability of certain chemicals of giving a horse an advantage, or being disadvantaged, in a race, contrary to the horse's inherent merits. To establish whether a prohibited substance is present, samples

(urine or blood, but can also be any part, or in contact with any part, of the horse) are taken from horses that have run in a race. Authorities may also take samples at any other time, according to their own specific rules. This varies from one country to another.

Sampling and forensic integrity

Any sample shall be collected under strict and secure chain of custody, with splitting into sample A and B. If the A sample is reported to contain prohibited substances, the B sample is usually analysed for those substances, either automatically or optionally at the trainer's or owner's request. This is generally done in the same place or in another Association of Official Racing Chemists (AORC) accredited laboratory.

Prohibited Substances

Officially, any xenobiotics are forbidden: horses are allowed only pure water, grass and oats. Exceptions are vitamins and a restricted list of antiparasitic and antimicrobial agents. Thus the lists for these animals are much larger than for humans and are usually given classes responsible for specific therapeutic actions. New additions and details are set by the Advisory Council to the International Federation of Horseracing Authorities on Doping Control.

The following are prohibited substances according to the current FEI veterinary regulations (effective 1 January 1998):

- substances, originating externally, whether they are endogenous or not to the horse and capable at any time of acting on one or more of the following mammalian body systems:

  nervous system,
  cardiovascular system,
  respiratory system,
  digestive system,
  urinary system,
  reproductive system,
  musculoskeletal system,
  skin (e.g. hypersensitizing agents),
  blood system,
  immune system, other than those in licensed vaccines,
  endocrine system, endocrine secretions and their synthetic counterparts;

- antipyretics, analgesics and antiinflammatory substances;
- cytotoxic substances.

This includes also any masking agents. Maximum threshold levels or ratios have also been established for those substances that are either endogenous or originate from common natural feed sources (see below).

As can be realized, there are quite a large number of substances that have to be tracked during screening and identified during confirmation procedures. It is impossible here to give all compounds that can potentially modify horse performance, but we will give a few examples of those that have been detected and reported in 1998 as the main substances:

16α-hydroxystanozolol
acepromazine and 2-(1-hydroxyethyl)-promazine
phenylbutazone, hydroxyphenylbutazone
oxyphenbutazone
benzydamine metabolite
boldenone
caffeine, theobromine, theophylline, paraxanthine
chlorobutanol
clanobutin
clenbuterol
dexamethazone
diclofenac
dimethylsulfoxide
ephedrine, norephedrine
nortestosterone
fentanyl, desethylfentanyl
flunixin
furosemide
isoxsuprine
ketoprofen, dihydroketoprofen
lidocaine, 3-hydroxylignocaine, lignocaine
methandriol
methocarbamol
methoxytyramine
naproxen
nefopam
nordiazepam
orphenadrine, nororphenadrine
polyethylene glycol
procaine
propantheline, 9-xanthenoic acid
pyrilamine
romifidine
salicylic acid
strychnine
testosterone
thioridazine, sulforidazine.

Regulatory identification

Different tests are used for different drugs or medicines. A broad range of tests is applied to every sample. Various types of target or special tests are used to screen for drugs and medicines that are not detected in broad-spectrum screening. Target tests are

used on a random basis. A finding of a prohibited substance means a finding of the substance itself or a metabolite of the substance, or an isomer of the substance or an isomer of a metabolite.

Regulatory quantification

Thresholds can only be adopted for:

- substances endogenous to the horse;
- substances arising from plants traditionally grazed or harvested as equine feed;
- substances in equine feed arising from contamination during cultivation, processing or treatment, storage or transportation.

For any finding of a prohibited substance of endogenous nature, the authorities may decide, either themselves or at the owner's or trainer's request, to examine the horse further.

Laboratory Performance Specification of the International Federation of Horseracing Authorities (IFHA)

AORC laboratories

The aim of future agreements between countries submitting samples to doping control analyses is such that the requested laboratory should present a minimum requirement for quality and performance:

- The laboratory should be accredited according to the current ISO 17025 regulations (previously called ISO guide 25), *General requirements for the competence of testing and calibration laboratories*, and to the supplementary document ILAC-G7, *Accreditation requirements and operating criteria for horseracing laboratories*.
- The laboratory should conform to the *Guide for establishing the presence of prohibited substances* (which is part B of ILAC-G7).
- The laboratory should meet the *Performance specification of the International Federation of Horseracing Authorities* (which is part C of ILAC-G7).
- The laboratory shall take part in interlaboratory comparisons (clause 5.9(b) of ISO 17025:1998).

AORC laboratories publish annual summary reports; these show that, in 1997, there were more than 600 000 analyses of saliva, urine and blood samples taken at equine events, with almost half of them performed in the USA. Saliva sampling seems to be mostly restricted to Illinois; and there are twice as many urine samples as blood samples. The number of positive cases is very few, 0.5% on average, with the highest rate in Europe.

Each year, proficiency testing is organized, with each laboratory having to analyse an average of eight samples with analyte concentrations ranging between 0.05 and $2\,\mu g\,l^{-1}$. It is not easy to know which techniques have been used, and little information is available. As a rule, each laboratory has to choose its own combination of techniques to achieve the best results over the number of samples submitted to it. In 1997, 63 different laboratories participated in the AORC proficiency test, which was mailed during the second week of March. Reports were received from 44 laboratories. The detection score was excellent, with the exception of bumetanide (a diuretic) and oxyphenbutazone. General considerations can, however, be given from the authors' own practical experiences in horse doping detection and other related fields like 'general unknown analyses' in forensic toxicology and human doping analyses.

- Each laboratory performs 1–5 (average 2) extraction procedures on the urine samples, prior to the use of one or more chromatographic techniques to detect various group of drugs and drug metabolites.
- A total of 14 different extraction procedures can be identified. The total volume of urine used for the extractions average 11 ml, and range from 3 to 20 ml.
- The base hydrolysis/acid extraction procedure is frequently used, followed by the cation exchange/reversed phase extraction. The dioctylsulfosuccinate paired ion extraction procedure seems to be no longer in use.
- The most frequently used chromatographic technique is TLC, followed by HPLC/DAD (photodiode array detector), specially in the USA. Over the years, it appears that there has been a decrease in the use of TLC and an increase in the use of HPLC/DAD and GC/MS.
- When using TLC, mostly on Silica Gel GF 254, several developing solvents combinations are used, together with various visualization reagents, noting that multiple visualizations are not necessarily overlaid on the same TLC plate. In many cases, more than one TLC system is used on the same urine extract, with different visualization techniques used on each plate.
- In HPLC/DAD screening, reverse phase systems with aqueous acid/organic mobile phases and solvent programming is used to analyse acidic extracts for antiinflammatory drugs. Computer programs allow reviewing of the data with respect to their respective retention and ultraviolet spectra data.
- A few laboratories make use of GC/NPD (nitrogen–phosphorus detector) analysis, mostly on basic

extracts with no derivatization. The analyses are performed on capillary columns with temperature programming. Retention times are evaluated manually and stimulants are mostly detected.
- Automated GC-MS screening is made possible with instruments equipped with suitable automatic injection devices and comparison with adequate reference mass spectra databases. Dedicated collections of MS are available from various commercial sources, but none is specifically dedicated to doping, except that of the AORC Drug Library, the 1998 revision of which contains 2104 entries.
- Screening urine samples with immunoassays is quite common and laboratories may rotate the use of up to 55 different drug immunoassays. These are always used in combination with at least one other chromatographic technique.

Special Requirements for Specific Compounds

The concept of threshold values for endogenous substances and substances of dietary origin is now well established in the field of equine drug testing. In order to be legally defensible, and to establish effective but fair means of control, it was imperative that these threshold values be introduced on a sound scientific basis. The corresponding research effort to achieve this objective has been met by a limited number of laboratories. Today, an agreement has been obtained on a 'no effect threshold'. Implicit in this approach is the requirement that these threshold values reflect drug concentrations that are pharmacologically inactive (**Table 1**). Substances below the thresholds are not actionable.

A few medications, for example phenylbutazone, are considered as permissible in some countries, depending on their regulations.

Other compounds might originate from natural sources or contamination of feeds and have resulted in controversy. For example, the growing practice of feeding manufactured compound feeds to horses in training has resulted in the detection of substances such as caffeine, theobromine, hordenine, lupanine, etc. in horse body fluids. The presence of these substances in horse feeds could be attributed to the choice of ingredients used in compounding the feed, or from crosscontamination on the production line. The philosophy of some national regulatory agencies in setting the threshold level for some of these substances (i.e. caffeine at 10 ng ml^{-1} plasma) is to overcome the problem of contaminated feeds causing 'unnecessary' withdrawal of horses from racing, with low, probably insignificant level of xenobiotics.

Table 1 Threshold values below which substances are considered to be inactive in horse doping testing

Substance	*'No effect threshold'*
Arsenic	0.3 μg total arsenic per ml in urine
Carbon dioxide	13 mmol available carbon dioxide per liter in plasma
Dimethyl sulfoxide	15 μg dimethyl sulfoxide per ml in urine, or 1 μg in plasma
Hydrocortisone	1 μg hydrocortisone per ml in urine
Nandrolone	Free and conjugated 5α-estrane-3β,17α-diol to free and conjugated 5(10)-estrane-3β,17α-diol in urine at a ratio of 1:1
Salicyclic acid	750 μg salicyclic acid per ml in urine, or 6.5 μg per ml in plasma
Testosterone	0.02 μg free and conjugated testosterone per ml in urine from geldings, or free and conjugated testosterone to free and conjugated epitesterone in urine from fillies and mares at a ratio of 12:1
Theobromine	2 μg per ml in urine

Elimination Rates for Xenobiotics

Published detection periods for interpretation by the veterinarian allow the estimation for safe withdrawal times of a treated horse with minimal risk of false positives. It is important that such information is based on sound scientific grounds.

The main difficulties in detecting doping agents in horses lies in the lack of reliable sources for pharmacokinetic data and the unknown structure of expected metabolites. Usually, each laboratory has its own source of positive urine and blood samples taken from horses that have received a low dose of one or several chemicals under controlled conditions. General commercial sources are those such as Sigma, Upjohn and Radian, for example. More specific sources can now provide drug standards for metabolites in calibrated solutions: Neogen Corp, Lexington, KY, USA and Gluck Equine Research Centre at the University of Kentucky, also in the USA.

In 1980, the Canadian Pari-Mutuel Agency (CPMA) initiated a research project, which had the following goals:

- to assess analytical methods in current use and determine their efficacy;
- to broaden the analytical coverage of prohibited drugs;
- to study the elimination and metabolism of 125–150 drugs in horses.

Much of the invaluable data developed in the CPMA research program are now published and serve as reference for many controlling laboratories. These reports provide methodology for the analysis of

individual drugs or groups of related drugs in equine body fluids, and also provide background information for those wishing to perform additional studies on the behavior of drugs administered to horses.

Conclusion

Drug testing in equine sport is a complex and difficult task for the analyst. The number of substances to be characterized is large, and the most modern instrumentation must be used for ultimate sensitivity and specificity. Since many samples prove to be negative, very efficient initial screening procedures are chosen in order to concentrate the laboratory's remaining facilities and resources on confirmation. Analysts in this field of forensic science, as in all others, need a high level of training and skill.

See also: **Analytical Techniques**: Presumptive Chemical Tests; Mass Spectrometry; Gas Chromatography; Spectroscopic Techniques; Hyphenated Chromatographic-Spectroscopic Techniques. **Toxicology**: Overview; Methods of Analysis – Ante Mortem. **Criminal Profiling**.

Further Reading

Fédération Equestre Internationale (1998) *Veterinary Regulations*, 8th edn. Lausanne: FEI.

Kallings P, Bondesson U and Houghton E (1995) *Proceedings of the 10th International Conference of Racing Analysts and Veterinarians*. Stockholm, 1994. Newmarket: R & W Publications.

Rivier L and Saugy M (1995) Systematic drug identification. In: A Townshend (ed.) *Encyclopaedia of Analytical Science*, pp. 1594–1606. London: Academic Press

Stevenson AJ, Weber MP, Hopkins D and Kacew S (1997) *Analytical Methodology for Detection and Confirmation of Drugs in Equine Body Fluids*. Ottawa: Canadian Pari-Mutuel Agency.

Inhalants

R J Flanagan, Guy's and St Thomas' Hospital Trust, London, UK

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0601

Introduction

If anesthesia is excluded, acute poisoning with volatile substances usually follows the deliberate inhalation of vapour in order to become intoxicated ('glue sniffing', inhalant abuse, solvent abuse, volatile substance abuse (VSA)). Solvents from adhesives, notably toluene, some correcting fluids and thinners (until recently, often 1,1,1-trichloroethane), hydrocarbons such as those found in cigarette lighter refills (usually liquefied petroleum gas (LPG)), aerosol propellants, halocarbon fire extinguishers and anesthetic gases such as nitrous oxide are among the compounds/products that may be abused in this way (**Tables 1** and **2**).

LPG arises from the cracking of oil to make petrol (gasoline) and also in the gases trapped above oil fields. There are at least two grades of LPG available to volatile substance abusers: (1) unpurified gas intended for direct use as a fuel; and (2) purified gas intended primarily for cigarette lighter refills and as a propellant in aerosols and related products. The composition of LPG can vary depending on the source, although its major components are usually butane, isobutane and propane. Some unpurified LPGs can contain up to 40% (v/v) unsaturates (butenes and propene).

Since the mid-1970s, concern as to the consequences of the release of volatile organochlorine and organobromine compounds such as chlorofluorocarbon (CFC) refrigerants and aerosol propellants into the atmosphere has led to the phased withdrawal of many such compounds in some countries. Deodorized LPG and dimethyl ether (DME), which is often used as a nonflammable azeotrope with a (chloro)-fluorocarbon, have already largely replaced fully halogenated CFCs as aerosol propellants in some parts of the world. In the case of refrigerants, the move is to polyfluorinated compounds such as difluoromethane, pentafluoroethane, 1,1,1,2-tetrafluoroethane, and 1,1,1-trifluoroethane (alone or as mixtures with other fluorocarbons) and, to a much lesser extent, perfluoropropane. Aliphatic hydrocarbons are also used to some extent as refrigerants. In addition, 1,1-dichloro-1-fluoroethane has been introduced as a degreasing and foam-blowing agent, and 1,1-difluoroethane and 1,1,1,2-tetrafluoroethane are sometimes used as aerosol propellants, either alone or as mixtures with other compounds.

It has often been stated that nail varnish or varnish remover (acetone and esters) may be abused by inhalation. However, properly documented examples of this practice have not come to light and it is probable that these compounds, although strong smelling, are too water-soluble to be intoxicants. For the same reason, acetone has never been used as an inhalational anesthetic. Diesel fuel, aviation fuel (kerosene, Avgas), white spirit, turpentine (or substitute) and paraffin are not sufficiently volatile to be abused by inhalation. Petrol (gasoline), on the other hand, is

Table 1 Some volatile substances which may be abused by inhalation

Hydrocarbons	Pentafluoroethane (FC 125)
Aliphatic	Perfluoropropane (octafluoropropane, FC 218)
Acetylene	Tetrachloroethylene (perchloroethylene)
Butane[a]	1,1,1,2-Tetrafluoroethane (FC 134a)
Isobutane (2-methylpropane)[a]	1,1,1-Trichloroethane (methylchloroform, Genklene)
Hexane[b]	1,1,1-Trifluoroethane (FC 143a)
Propane[a]	1,1,2-Trichlorotrifluoroethane (FC 113)
Alicyclic/aromatic	Trichloroethylene ('trike', Trilene)
Cyclopropane (trimethylene)	Trichlorofluoromethane (FC 11, Freon 11)
Toluene (toluol, methylbenzene, phenylmethane)	
Xylene (xylol, dimethylbenzene)[c]	**Oxygenated compounds and others**
Mixed	Butanone (2-butanone, methyl ethyl ketone, MEK)
Petrol (gasoline)[d]	Butyl nitrite[f]
Petroleum ethers[e]	Enflurane [(R,S)-2-chloro-1,1,2-trifluoroethyl difluoromethyl ether]
Halogenated	Ethyl acetate
Bromochlorodifluoromethane (BCF, FC 12B1)	Desflurane [(R,S)-difluoromethyl 1,2,2,2-tetrafluoroethyl ether]
Carbon tetrachloride (tetrachloromethane)	Diethyl ether (ethoxyethane)
Chlorodifluoromethane (FC 22, Freon 22)	Dimethyl ether (DME, methoxymethane)
Chloroform (trichloromethane)	Isobutyl nitrite ('butyl nitrite')[f]
Dichlorodifluoromethane (FC 12, Freon 12)	Isoflurane [(R,S)-1-chloro-2,2,2-trifluoroethyl difluoromethyl ether]
1,1-Dichloro-1-fluoroethane (FC 141b, Genetron 141b)	Isopentyl nitrite (3-methylbutan-1-ol, isoamyl nitrite, 'amyl nitrite')[f, g]
Dichloromethane (methylene chloride)	Methoxyflurane (2,2-dichloro-1,1-difluoroethyl methyl ether)
1,2-Dichloropropane (propylene dichloride)	Methyl acetate
1,1-Difluoroethane (FC 152a)	Methyl isobutyl ketone (MIBK, isopropyl acetone, 4-methyl-2-pentanone)
Difluoromethane (FC 32)	Methyl tert.-butyl ether (MTBE)
Ethyl chloride (monochloroethane)	Nitrous oxide (dinitrogen monoxide, 'laughing gas')
Halothane [(R,S)-2-bromo-2-chloro-1,1,1-trifluoroethane]	Sevoflurane [fluoromethyl 2,2,2-trifluoro-1-(trifluoromethyl)ethyl ether]
	Xenon

[a] Principal components of purified liquified petroleum gas (LPG).
[b] Commercial 'hexane' mixture of hexane and heptane with small amounts of higher aliphatic hydrocarbons.
[c] Mainly *meta*-xylene (1,3-dimethylbenzene).
[d] Mixture of aliphatic and aromatic hydrocarbons with boiling range 40–200°C.
[e] Mixtures of pentanes, hexanes, etc. with specified boiling ranges (e.g. 40–60°C).
[f] Abused primarily for its vasodilator properties.
[g] Commercial 'amyl nitrite' mainly isopentyl nitrite but other nitrites also present.

often abused, especially in less developed communities. Isobutyl and isopentyl ('amyl') nitrites are also inhaled in order to experience their vasodilator properties, sometimes by male homosexuals. In addition, those who ingest, or even more rarely inject, solvents or solvent-containing products, either accidentally or deliberately, and the victims of clinical, industrial and domestic accidents may be poisoned by the compounds under consideration. Finally, chloroform, diethyl ether and other volatiles are still used occasionally in the course of crimes such as rape and murder, while a further volatile compound, chlorobutanol (1,1,1-trichloro-2-methyl-2-propanol), sometimes employed as a sedative and a preservative, has been used in doping racing greyhounds.

Diagnosis of Acute Poisoning with Volatile Substances

VSA should be suspected in children and adolescents with 'drunken' behavior, unexplained listlessness, anorexia and moodiness. The hair, breath and clothing may smell of solvent, and empty adhesive tubes or other containers, potato crisp bags, cigarette lighter refills or aerosol spray cans are often found. The smell of solvent on the breath is related to the dose and duration of exposure and may last for many hours. The so-called 'glue-sniffer's rash' (perioral eczema) is probably caused by repeated contact with glue in a plastic or other bag held to the face. Although primarily a phenomenon of adolescence, it must be remembered that adults, especially those such as dentists, operating theatre personnel, and medical and dental students, with ready occupational access to abusable volatile compounds, also indulge in VSA. In the late 1970s, for example, it was estimated that some 1–1.6% of US dentists were abusing nitrous oxide.

The major risk from VSA is the possibility of sudden death. There have been at least 1596 (1394 male) VSA-related deaths ('sudden sniffing deaths') in the UK in the period 1971–1996 (**Fig. 1**), some

Table 2 Some products which may be abused by inhalation

Product	*Major volatile components*
Adhesives	
Balsa wood cement	Ethyl acetate
Contact adhesives	Butanone, hexane, toluene and esters
Cycle tyre repair cement	Toluene and xylenes
Polyvinylchloride (PVC) cement	Acetone, butanone, cyclohexanone, trichloroethylene
Woodworking adhesives	Xylenes
Aerosols	
Air freshener	LPG, DME and/or fluorocarbons[a]
Deodorants, antiperspirants	LPG, DME and/or fluorocarbons[a]
Fly spray	LPG, DME and/or fluorocarbons[a]
Hair lacquer	LPG, DME and/or fluorocarbons[a]
Paint	LPG, DME and/or fluorocarbons[a] and esters
Anesthetics/analgesics	
Inhalational	Nitrous oxide, cyclopropane, diethyl ether, halothane, enflurane, desflurane, isoflurane, methoxyflurane, sevoflurane, xenon
Topical	Ethyl chloride, fluorocarbons[a]
Dust removers ('air brushes')	DME, fluorocarbons[a]
Commercial dry cleaning and degreasing agents	Dchloromethane, FC 113, FC 141b, methanol, 1,1,1-trichloroethane, tetrachloroethylene, toluene, trichloroethylene (now very rarely carbon tetrachloride, 1,2-dichloropropane)
Domestic spot removers and dry cleaners	Dichloromethane, 1,1,1-trichloroethane, tetrachloroethylene, trichloroethylene
Fire extinguishers	BCF, FC 11, FC 12
Fuel gases	
Cigarette lighter refills	LPG
'Butane'	LPG
'Propane'	Propane and butanes
Paints/paint thinners	Acetone, butanone, esters, hexane, toluene, trichloroethylene, xylenes
Paint stripper	Dichloromethane, methanol, toluene
Racing fuel super-charge tanks	Nitrous oxide
'Room odorizer'	Isobutyl nitrite
Surgical plaster/chewing gum remover	1,1,1-Trichloroethane, trichloroethylene
Typewriter correction fluids/thinners (some)	1,1,1-Trichloroethane
Whipped cream dispensers	Nitrous oxide

[a] Mainly FC 134a.
See Table 1 for full chemical names of some compounds.

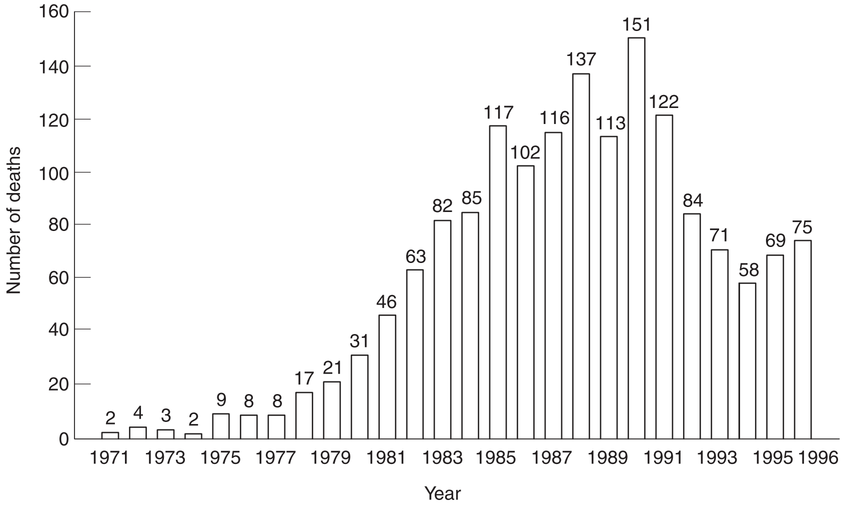


Figure 1 UK VSA-related deaths, 1971–1996 (n = 1596). (Source: Taylor *et al*., 1998.)

60% in adolescents aged 14–18 years. There have also been many reports of VSA-related sudden deaths in other parts of the world. Some deaths occur from 'indirect' causes such as inhalation of vomit (14.8% of UK VSA-related deaths), asphyxia associated with the use of a plastic bag to contain the vapor being inhaled (11.5%), and trauma (11.7%). However, most UK deaths (53.3%) were attributed to 'direct toxicity'. These effects could include anoxia, respiratory depression or vagal stimulation, all leading to cardiac arrest. Alternatively, direct cardiotoxicity is a possibility. Whatever the precise sequence of events leading to death, it is clear that 'direct toxicity' is reported more commonly in deaths due to abuse of fuel gases, aerosol propellants and halogenated solvents than in deaths due to abuse of solvents from adhesives (**Fig. 2**).

The analytical toxicology laboratory may be asked to perform analyses for solvents and other volatile compounds in biological samples and related specimens to:

- assist in the diagnosis of acute poisoning;
- confirm a suspicion of chronic VSA in the face of denial from the patient and/or a caretaker;
- aid the investigation of deaths where poisoning by volatile compounds is a possibility, including deaths associated with anesthesia;
- aid investigation of rape or other assault, or other offence such as driving a motor vehicle or operating machinery, which may have been committed under the influence of volatile substances;
- aid investigation of incidents such as rape or other assault in which volatile substances may have been administered to the victim;
- help investigate fire or explosion where VSA might have been a contributory factor; and
- assess occupational or environmental exposure to anesthetic or solvent vapor. However, other techniques such as ambient air monitoring or, in a few instances, the measurement of urinary metabolite excretion may be more appropriate in this latter context.

The laboratory analysis of volatile substances presents particular problems. Firstly, many of the compounds of interest occur commonly in laboratories and this necessitates special precautions against contamination and interference. Secondly, collection, storage and transport of biological samples must be controlled as far as practicable in order to minimize loss of analyte: quantitative work is futile if very volatile compounds such as propane are encountered unless special precautions are taken to prevent the loss of analyte from the sample prior to the analysis. Thirdly, many compounds of interest are excreted unchanged via the lungs and thus blood (and/or other tissues in fatalities), and not urine, is usually the sample of choice. Finally, the interpretation of results can be difficult, especially if legitimate exposure to solvent vapor is a possibility.

A diagnosis of VSA should be based on a combination of circumstantial, clinical and analytical evidence rather than on any one factor alone. It is especially important to consider all circumstantial evidence in cases of possible VSA-related sudden death, as suicide or even homicide cannot be excluded simply on the basis of the toxicological examination. There have been a number of reports of the use of inhalational anesthetics for suicidal purposes, for

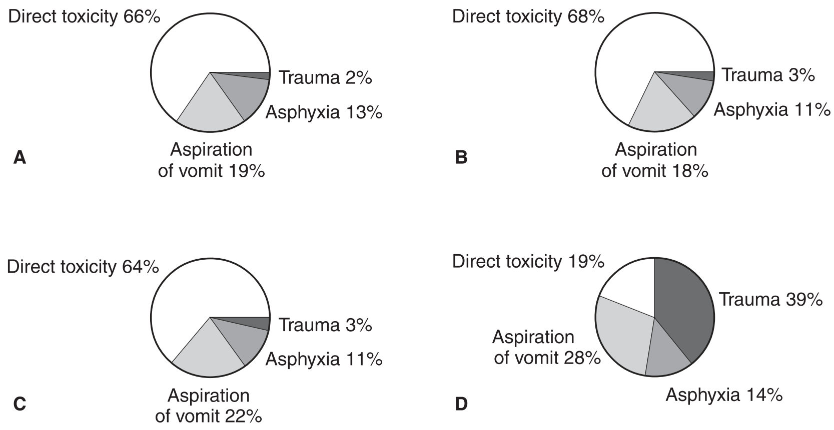


Figure 2 UK VSA-related deaths by mechanism and by type of product abused. (Source Taylor *et al.*, 1998.) (**A**) Aerosol propellants; (**B**) fuel gases; (**C**) chlorinated and other solvents; (**D**) solvents from adhesives.

example, and there has been one example in the UK of a serial murderer whose victims were thought initially to have died as a result of VSA.

Analytical Methods

The analysis of biological samples for solvents and other volatiles which may be abused by inhalation has similarities to the analysis of methanol, ethanol and 2-propanol. However, poisoning with these latter compounds is normally the result of ingestion or occupational exposure to vapour. Gas chromatography (GC) with flame ionization and/or electron capture detection (FID and/or ECD) is widely used in the analysis of volatiles in blood and other biological specimens which may be obtained without using special apparatus such as breath-collection tubes. Nitrous oxide and most halogenated compounds respond on the ECD, although the thermal conductivity detector (TCD) may be used as an alternative if nitrous oxide poisoning is suspected. Although not reported as yet, abuse of xenon is a possibility. GC methods for this element would require TCD or mass spectrometry (MS). Direct MS of expired air can also detect many compounds several days postexposure. However, at present the use of this technique is limited by the need to take breath directly from the patient. Vapour-phase infrared (IR) spectrophotometry may be useful in the analysis of abused products or ambient atmospheres.

If the analyte is very volatile (e.g. propane or butane) and a quantitative analysis is required, a blood sample should be collected directly into the headspace vial in which the analysis will be carried out. Many other volatile compounds are relatively stable in blood and other tissues if simple precautions are taken. In the case of blood, the container used for the sample should be glass, preferably with a cap lined with metal foil; greater losses may occur if plastic containers are used. The tube should be as full as possible and should only be opened when required for analysis, and then only when cold (4°C). If the sample volume is limited it is advisable to select the container to match the volume of blood so that there is minimal headspace. An anticoagulant (sodium ethylenediamine tetra-acetate (EDTA) or lithium heparin) should be used. Specimen storage between −5 and 4°C is recommended and 1% (w/v) sodium fluoride should be added to minimize esterase and other enzymic activity. Tissues (approximately 10 g each of brain, lung, liver, kidney and subcutaneous fat) should also be obtained, if a necropsy is to be performed, in addition to standard toxicological specimens if available (femoral blood, urine, stomach contents, vitreous humor). Tissues should be stored before analysis in the same way as blood. No preservative should be added. Products thought to have been abused or otherwise implicated in the incident (and stomach contents if ingestion is suspected) should be packed, transported and stored entirely separately from (other) biological specimens to avoid crosscontamination. Investigation of deaths occurring during or shortly after anesthesia should include the analysis of the inhalation anesthetic(s) used in order to exclude an administration error.

Gas chromatography

A summary of GC methods for volatile compounds in biological specimens published since 1989 is given in **Table 3**. Packed columns, for example 2 m × 2 mm internal diameter (i.d.) 0.3% (w/w) Carbowax 20M on Carbopack C programmed from 35 to 175°C, have been used extensively in conjunction with headspace sample preparation. On-column septum injections of up to 400 μl headspace can be performed and good sensitivity (of the order of 0.1 mg l^{-1} or better using 200 μl of sample) can be obtained. Disadvantages include the poor resolution of some very volatile substances, a long total analysis time, and variation in the peak shape given by alcohols between different batches of column packing. Porous layer open tubular (PLOT) columns give good retention, and thus resolution, of compounds with similar relative formula mass, but peak shapes of polar compounds are poor and it is difficult to screen for compounds of widely different volatility in one analysis.

Bonded-phase wide-bore capillary columns permit relatively large volume septum injections and can offer advantages of improved efficiency, reproducibility and reliability. A 60 m × 0.53 mm i.d. fused silica capillary coated with the dimethylpolysiloxane SPB-1 (5 μm film thickness) programmed from 40 to 200°C offers many advantages over packed column and PLOT systems. Improved resolution of very volatile compounds is obtained and, even with an initial temperature of 40°C, the total analysis time can be reduced to 26 min, and good peak shapes can be obtained even for alcohols (**Fig. 3**). Moreover, splitless septum injections of up to 300 μl headspace can be performed with no noticeable effect on column efficiency; hence, sensitivity is as least as good as that attainable with a packed column.

The use of a capillary column together with two different detectors (FID and ECD) confers a high degree of selectivity, particularly for low formula mass compounds where there are very few alternative structures. If more rigorous identification is required,

Table 3 Summary of gas chromatographic methods for volatile compounds in biological fluids published 1989–1997

Application	*Sample*	*Extraction*[a]	*GC column and/or conditions*	*Detection*	LoD[b]
Anesthetics	Blood (0.6 ml)	HS	15% Apiezon L, 80/100 mesh Chromosorb W, 2 m × 4 mm i.d., 130°C	FID	
Toxicology	Blood, tissue	HS	DB-1701, 40 m × 0.25 mm i.d., 1.0 μm film, 30–120°C	MS (ion trap)	
Toxicology	Blood, tissue (1 g)	HS	15% Carbowax 1500, 80/100 mesh Chromosorb W NAW, 1.8 m × 0.32 mm i.d., 70°C	FID	
Toxicology	Blood, tissue	HS	30% Carbowax 20 M, 60/80 mesh Chromosorb W AW-DMCS, 2 m, 70°C	FID	
Anesthetics	Blood (1 ml)	HS	Porapak S, 80/100 mesh, 1.9 m × 2 mm i.d., 165°C	FID	
Environmental	Blood (10 ml)	PT	DB-624, 30 m × 0.32 mm i.d., 1.8 μm film, 0–190°C	MS	$0.05\ \mu g\ l^{-1}$[c]
Toxicology	Blood (0.2 ml), tissue	HS	SPB-1, 60 m × 0.32 mm i.d., 5 μm film, 40–200°C	FID, ECD	
Toxicology	Blood (1 ml), tissue (1 g)	HS	0.2% Carbowax 1500, 80/100 mesh Carbopak C, 1.8 m, 100°C	FID	
Toxicology	Blood (1.5 ml)	HS	DB-1, 30 m × 0.25 mm i.d., 1.0 μm film, 40–250°C	MS (ion trap)	
Workplace monitoring	Urine (0.5–1 ml)	HS	Porapak Q, 80–100 mesh, 2 m × 1.8 mm i.d., 100°C	MS	$0.7\ \mu g\ l^{-1}$[d]
			HP-5, 25 m × 0.2 mm i.d., 25°C	MS	$0.1\ \mu g\ l^{-1}$
Toxicology	Blood, CSF (1 ml)	HS	Porapak Q, 80/100 mesh, 1 m × 2.5 mm i.d., 50–170°C	MS	$0.02\ mg\ l^{-1}$[e]
Workplace monitoring	Urine (50 ml)	HS	DB-5, 30 m × 0.25 mm i.d., 40°C	MS	
Toxicology	Blood, plasma, serum (1–3 μl)	PH	GS-Q, 30 m × 0.53 mm i.d., 70–230°C (screening)	MS	
		PH	Porapak Q, 1 m × 2.6 mm i.d., 180°C (screening)	FID	
		PH	DB-17, 15 m × 0.53 mm i.d., 1.0 μm film, 50°C (quantification)	MS	
Anesthetics	Blood (3 μl)	PH	GS-Q, 30 m × 0.53 mm i.d., 160°C	MS	$0.2\ mg\ l^{-1}$
	Blood (0.5 ml)	HS	DB-1, 30 m × 0.53 mm i.d., 5 μm film, 60°C	MS	
Toxicology	Blood (0.5–5 ml)	PT	PoraPLOT Q, 25 m × 0.32 mm i.d., 10 μm film, 30–250°C	FTIR, FID	$0.05\ mg\ l^{-1}$[e]
Toxicology	Blood, brain (35 mg)	HS-FE	DB-1, 30 m × 0.25 mm i.d., 1.0 μm film, 40–250°C	MS (ion trap)	

[a] HS = headspace, HS-FE = headspace full evaporation, PH = pulse heating, PT = purge and trap
[b] LoD = Limit of detection (in some cases limit of identification or limit of accurate measurement)
[c] 1,1,1-Trichloroethane
[d] Nitrous oxide
[e] Isoflurane
Adapted from Pihlainen and Ojanperä (1998).

GC combined with MS or Fourier transform IR spectrometry (FTIR) may be used. However, GC-MS can be difficult when the fragments produced are less than m/z 40, particularly if the instrument is used for other purposes as well as solvent analyses. In particular, the available sensitivity and spectra of the low molecular weight alkanes renders them very difficult to confirm by GC-MS. Inertial spray MS allows introduction of biological fluids directly into the MS without prior chromatographic analysis and has been used in the analysis of halothane in blood during anesthesia.

GC-FTIR is generally more appropriate than GC-MS in the analysis of volatiles, but sensitivity is poor, particularly when compared with the ECD. In addition, interference, particularly from water and carbon

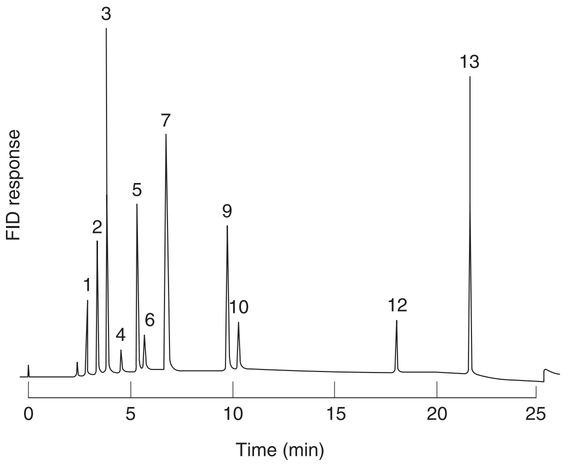


Figure 3 Analysis of a whole blood sample obtained post mortem from a patient who had inhaled the contents of a cigarette lighter refill. Sample preparation: internal standard solution (ethylbenzene and 1,1,2-trichloroethane (approximately 25 and 10 mg l^{-1}, respectively) in outdated blood-bank whole blood) (200 μl) incubated (65°C, 15 min) with specimen (200 μl) in a sealed 10 ml glass septum vial. Chromatographic conditions: column: 60 m × 0.53 mm i.d. SPB-1 (5 μm film). Oven temperature: 40°C (6 min), then to 80°C at 5°C min^{-1}, then to 200°C at 10°C min^{-1} Injection: 300 μl headspace. 1, Propane; 2, isobutane; 3, butane; 4, ethanol; 5, acetone; 6, 2-propanol; 7, 2-methyl-2-propanol; 9, butanone; 10, 2-butanol; 12, 1,1,2-trichloroethane (IS); 13, ethylbenzene (IS).

dioxide in the case of biological specimens, can be troublesome. 'Purge and trap' and multiple headspace extraction offer ways of increasing sensitivity and, although not needed for most clinical and forensic applications, have been used either in conjunction with GC-FTIR, or in occupational/environmental monitoring. Pulse heating has also been employed in the analysis of volatiles in biological specimens (Table 3). Advantages of this latter technique include use of a small sample volume (0.5–5 μl), short extraction time and lack of matrix effects.

Pharmacokinetics and the Interpretation of Results

Knowledge of the pharmacokinetics of volatile compounds is important in understanding the rate of onset, the intensity and the duration of intoxication with these substances. The UK maximum exposure limit (MEL) or occupational exposure standard (OES) (**Table 4**) provide information on the relative toxicities of different compounds after chronic exposure to relatively low concentrations of vapor. Inhaled compounds may rapidly attain high concentrations in well-perfused organs (brain, heart), whereas concentrations in muscle and adipose tissue may be very low. Should death occur, this situation is 'frozen' to an extent, but, if exposure continues, the compound will accumulate in less accessible (poorly perfused) tissues, only to be slowly released once exposure ceases. Thus, the plasma concentrations of some compounds may fall monoexponentially, while others may exhibit two (or more) separate rates of decline (half-lives).

The solubility of a volatile compound in blood is an important influence on the rate of absorption, tissue distribution and elimination of the compound. The partition coefficients of a number of compounds between air, blood and various tissues have been measured *in vitro* using animal tissues, and some *in vivo* distribution data have been obtained from postmortem tissue measurements in humans (Table 4). However, these latter data must be used with caution, as there are many difficulties inherent in such measurements (sampling variations, analyte stability, external calibration, etc.). Published data on the elimination half-lives of volatile substances (Table 4) are not easily comparable, either because too few samples were taken or the analytical methods used did not have sufficient sensitivity to measure the final half-life accurately.

Metabolism of volatile substances

Exogenous compounds may be metabolized in a number of ways, a frequent result being the production of metabolites of greater polarity (water solubility) and thus lower volatility than the parent compound. The pharmacological activity and pharmacokinetics of any metabolite(s) often differ from those of the parent compound(s). After ingestion, extensive hepatic metabolism can reduce systemic availability ('first-pass' metabolism). Many volatile substances, including butane, dimethyl ether, most fluorocarbon refrigerants/aerosol propellants, isobutane, nitrous oxide, propane, tetrachloroethylene and 1,1,1-trichloroethane, are largely eliminated unchanged in exhaled air. Others are partly eliminated in exhaled air and also metabolized in the liver and elsewhere, the metabolites being eliminated in exhaled air or in urine (**Table 5**), or incorporated into intermediary metabolism.

Interpretation of qualitative results

The likelihood of detecting exposure to volatile substances by headspace GC of blood is influenced by the dose and duration of exposure, the time of sampling in relation to the time elapsed since exposure, and the precautions taken in the collection and storage of the

Table 4 Physical properties and pharmacokinetic data of some volatile compounds

Compound	*MEL/OES ($mg\,m^{-3}$)*[a]	*Vapour pressure (20 °C, mmHg)*[b]	*Inhaled dose absorbed (%)*	*Proportion absorbed dose (%)*		*Half-life (h)*[c]	*Brain:blood distribution ratio (deaths)*	*Partition coefficient (blood:gas) (37 °C)*
				Eliminated unchanged	*Metabolized*			
Acetone	1810	183	–	–	–	3–5[d]	–	243–300
Benzene	16	75	46	12	80	9–24	3–6	6–9
Butane	1750[e]	(1554)	30–45	–	–	–	–	–
Isobutane	1750[e]	(2282)	–	–	–	–	–	–
Butanone	600	75	70	99+	0.1	0.5	–	116
Carbon disulfide	32	294	40	<30	50–90	< 1	–	2.4
Carbon tetrachloride	13	90	–	50?	50?	48	–	1.6
Chlorodifluoromethane	3590	(6701)	–	–	–	–	1.9	–
Chloroform	9.9	157	–	20–70 (8 h)	>30	–	4	8
Cyclopropane	–	(4701)	–	99	0.5	–	1.5–3.6	0.55
Desflurane	–	669	–	–	0.02	–	1.29[f]	0.42
Dichlorodifluoromethane	5030	(3639)	35	99	< 0.2	–	1.4	0.15
Dichloromethane	350	350	–	50?	<40	0.7	0.5–1	5–10
Diethyl ether	1230	438	–	>90	–	–	1.1	12
Enflurane	383	172	90+	>80 (5 days)	2.5	36	1.4[e]	1.9
Ethyl acetate	1460	72	–	–	–	–	–	–
Halothane	82	244	90+	60–80 (24 h)	<20	2–3	2–3	2.57
Hexane	72	122	–	–	–	–	–	–
Isoflurane	383	240	–	–	0.2	–	1.57[f]	1.38
Methoxyflurane	–	23	–	19 (10 days)	>44	–	2–3	13
Methyl isobutyl ketone	208	15	–	–	–	–	–	–
Nitrous oxide	183	(39 800)	–	>99	–	–	1.1	0.47
Propane	1750[e]	(6269)	–	–	–	–	–	–
Sevoflurane	–	157	–	–	3	20	1.7[f]	0.68
Styrene	430	4	–	1–2	>95	13	–	32
Tetrachloroethylene	345	14	60+	>90	1–2	72	9–15	9–19
Toluene	50	22	53	<20	80	7.5	1–2	8–16
1,1,1-Trichloroethane	1110	98	–	60–80 (1 week)	2	10–12	2	1–3
Trichloroethylene	550	58	50–65	16	>80	30–38	2	9.0
Trichlorofluoromethane	5710	667	92	89	< 0.2	1.5	2.5	0.87
'Xylene'	441	6	64	5	>90	20–30	–	42.1

[a] UK maximum exposure limit/occupational exposure standard (8 h time weighted average) 1998.
[b] Figures in parentheses indicate compound gas at 20°C.
[c] Terminal phase elimination half-life.
[d] Longer after high doses.
[e] As components of liquified petroleum gas (LPG).
[f] Experimental: 37°C.
Data summarized from Fiserova-Bergerova (1983), Moffatt *et al.* (1986), Baselt and Cravey (1995) and Pihlainen and Ojanperä (1998).

Table 5 Summary of the metabolism of some solvents and other volatile substances

Compound	*Principal metabolites (% absorbed dose)*	*Notes*
Acetone	2-Propanol (minor) and intermediary metabolites (largely excreted unchanged at higher concentrations)	Endogenous compound produced in large amounts in diabetic or fasting ketoacidosis; also the major metabolite of 2-propanol in man
Acetonitrile	Inorganic cyanide (at least 12%) thence to thiocyanate	Cyanide/thiocyanate may accumulate during chronic exposure
Benzene	Phenol (51–87%), catechol (6%), hydroquinone (2%), *trans,trans*-muconic acid	Excreted in urine as sulfate and glucuronide conjugates. Urinary phenol excretion has been used to indicate exposure but is variable and subject to interference
Bromomethane	Inorganic bromide	Serum bromide has been used to monitor exposure, although the concentrations associated with toxicity are much lower than when bromide itself given orally
Butanone	3-Hydroxybutanone (0.1%)	3-Hydroxybutanone excreted in urine. Most of an absorbed dose of butanone excreted unchanged in exhaled air
Carbon disulfide	2-Mercapto-2-thiazolin-5-one, 2-thiothiazolidine-4-carboxylic acid (TCCA), thiourea, inorganic sulfate and others	2-Mercapto-2-thiazolin-5-one glycine conjugate and TCCA glutathione conjugate of carbon disulphide. Urinary TCCA excretion reliable indicator of exposure
Carbon tetrachloride	Chloroform, carbon dioxide, hexachloroethane and others	Trichloromethyl free radical (reactive intermediate) probably responsible for hepatorenal toxicity
Chloroform	Carbon dioxide (up to 50%), diglutathionyl dithiocarbonate	Phosgene (reactive intermediate) depletes glutathione and is probably responsible for hepatorenal toxicity
Cyclohexanone	Cyclohexanol, *trans*-1,2-cyclohexanediol, *trans*-1,4-cyclohexanediol	Metabolites excreted mainly as glucuronides in adults
Desflurane	Trifluoroacetic acid ($< 0.02\%$), inorganic fluoride	–
Dichloromethane	Carbon monoxide ($\pm 35\%$)	CO blood half-life 13 h breathing air (atmospheric pressure) (CO half-life 5 h after inhalation of CO itself). Blood carboxyhemoglobin measurement useful indicator of chronic exposure
Dimethylsulfoxide	Dimethylsulfide (3%), dimethylsulfone (18–22%)	After oral/dermal administration, dimethyl sulfide excreted in exhaled air and dimethylsulfone in urine
Dioxane	β-Hydroxyethoxyacetic acid (HEAA)	HEAA excreted in urine
Enflurane	Difluoromethoxydifluoroacetic acid ($> 2.5\%$), inorganic fluoride	–
Ethyl acetate	Ethanol, acetic acid	Rapid reaction catalyzed by plasma esterases
Ethylbenzene	Methylphenylcarbinol (5%), mandelic acid (64%), phenylglyoxylic acid (25%)	Methylphenylcarbinol excreted in urine as conjugate, others as free acids. Mandelic acid excretion has been used to monitor ethylbenzene exposure
Halothane	2-Chloro-1,1,1-trifluoroethane, 2-chloro-1,1-difluoroethylene, trifluoroacetic acid, inorganic bromide and others	The formation of reactive metabolites may be important in the etiology of the hepatotoxicity ('halothane hepatitis') which may occur in patients re-exposed to halothane or similar compounds
Hexane	2-Hexanol, 2-hexanone, 2,5-hexanedione	Hexan-2-ol excreted in urine as glucuronide. 2,5-Hexanedione thought to cause neurotoxicity. Methyl butyl ketone also neurotoxic and also metabolized to 2,5-hexanedione
Isobutyl nitrite	2-Methyl-1-propanol (99%+), inorganic nitrite	Parent compound not normally detectable in blood. Blood methemoglobin can be used to monitor exposure
Isoflurane	Trifluoroacetic acid ($< 0.2\%$), inorganic chloride, inorganic fluoride	–
Isopentyl nitrite	3-Methyl-1-butanol (99%+), inorganic nitrite	Parent compound not normally detectable in blood. Blood methemoglobin can be used to monitor exposure
Methanol	Formaldehyde (up to 60%), formic acid	Urinary formic acid excretion has been advocated for monitoring methanol exposure
Methoxyflurane	Methoxydifluoroacetic acid, dichloroacetic acid, oxalic acid, inorganic fluoride	Abuse has been manifested as chronic fluoride poisoning (fluorosis)
2-Propanol	Acetone (80–90%) thence others	2-Propanol blood half-life $\pm$ 2 h, acetone half-life $\pm$ 22 h.

Table 5 (continued)

Compound	*Principal metabolites (% absorbed dose)*	*Notes*
Sevoflurane	1,1,1,3,3,3-Hexafluoropropanol (<3%), inorganic fluoride	1,1,1,3,3,3-Hexafluoropropanol excreted as glucuronide in urine
Styrene	Mandelic acid (85%) and phenylglyoxylic acid (10%); hippuric acid may be minor metabolite	Urinary mandelic acid excretion indicates exposure. Ethanol inhibits mandelic acid excretion
Tetrachloroethylene	Trichloroacetic acid (< 3%)	Urinary trichloroacetic acid excretion serves only as qualitative index of exposure
Toluene	Benzoic acid (80%) and *ortho*-, *meta*- and *para*-cresol (1%)	Benzoic acid largely conjugated with glycine giving hippuric acid which is excreted in urine (blood half-life 2–3 h). Not ideal index of exposure since there are other (dietary) sources of benzoic acid
1,1,1-Trichloroethane	2,2,2-Trichloroethanol (2%) and trichloroacetic acid (0.5%)	Urinary metabolites serve as qualitative index of exposure only (compare tetrachloroethylene)
Trichloroethylene	2,2,2-Trichloroethanol (45%) and trichloroacetic acid (32%)	Trichloroethanol (glucuronide) and trichloroacetic acid excreted in urine (blood half-lives about 12 and 100 h respectively). Trichloroacetic acid excretion can indicate exposure
Xylenes	Methylbenzoic acids (95%) and xylenols (2%)	Methylbenzoic acids conjugated with glycine and urinary methylhippuric acid excretion used as index of exposure – no dietary sources of methylbenzoates.

specimen. In a suspected VSA- or anesthetic-related fatality, analysis of tissues (especially fatty tissues such as brain) may prove useful, as high concentrations of volatile compounds may be present even if very little is detectable in blood.

Analysis of metabolites in urine may extend the time in which exposure may be detected but, of the compounds commonly abused, only toluene, the xylenes and some chlorinated solvents, notably trichloroethylene, have suitable metabolites (Table 5). Chronic petrol 'sniffing' has been diagnosed by the measurement of blood lead concentrations or detection of aromatic components such as toluene and ethylbenzene. Abuse of the fluorinated anesthetic methoxyflurane has been detected by measuring serum and urine fluoride ion concentrations. With some petrols and other complex mixtures such as petroleum ethers (Table 1), however, the blood concentrations of the individual components are often below the limit of detection of headspace GC methods even after significant inhalational exposure.

Detection of a volatile compound in blood does not always indicate VSA or occupational/environmental exposure to solvent vapor. Acetone and some of its homologs may occur in high concentrations in ketotic patients. Large amounts of acetone and butanone may also occur in blood and urine from children with acetoacetylcoenzyme A thiolase deficiency, for example, and may indicate the diagnosis. In addition, acetone is the major metabolite of 2-propanol in humans (Table 5). Conversely, 2-propanol has been found in blood from ketotic patients. Other ketones may also give rise to alcohols *in vivo*. Cyclohexanol, for example, is the principal metabolite of cyclohexanone in humans (Table 5). Other volatile compounds, such as halothane or chlorobutanol, may be used in therapy or inadvertently added to the sample as a preservative. When interpreting the results of qualitative analyses it is important to remember that some compounds often occur in association one with another (**Table 6**).

Use of aerosol disinfectant preparations when collecting specimens may contaminate the sample if an aerosol propellant is used. Contamination of blood samples with ethanol or 2-propanol may also occur if an alcohol-soaked swab is used to cleanse skin prior to venepuncture. Gross contamination with technical xylene (a mixture of *o*-, *m*- and *p*-xylene together with ethylbenzene) has been found in blood collected into Sarstedt Monovette serum gel blood collection tubes; contamination with toluene (up to $22\,mg\,l^{-1}$), 1-butanol, ethylbenzene and xylene has been found in more recent batches of these same tubes. Contamination with 1-butanol or 2-methyl-2-propanol occurs commonly in blood collected into tubes coated with EDTA. Care should be taken when handling frozen tissue prior to analysis as any compounds present in ambient air may condense on the cold surface and

Table 6 Associated volatile compounds

Compound	*Associated compound(s)*
Acetone	Butanone and higher ketones in ketoacidosis, 2-propanol (metabolite, rare)
BCF	FC 11
Butane	Butanone (metabolite[a]), isobutane, 2-butanol (metabolite[a]), propane
Cyclohexanone	Cyclohexanol (metabolite)
Dimethyl ether	FC 22 or other fluorocarbon
Ethanol	Propanols and higher alcohols if bacterial fermentation has occurred; methanol or other volatile poisons if denatured alcohol has been consumed
Ethyl acetate[b]	Ethanol (metabolite)
Ethylbenzene	(see Xylenes below)
FC 11	BCF, FC 12
FC 12	FC 11
FC 22	Dimethyl ether
Halothane	2-Chloro-1,1-difluoroethylene, 2-chloro-1,1,1-trifluoroethane (metabolites[a])
Isobutane	Butane, 2-methyl-2-propanol (metabolite[a]), propane
Isobutyl nitrite[b]	2-Methyl-1-propanol (degradation product)
Isopentyl nitrite[b]	3-Methyl-1-butanol (degradation product)
Methyl acetate[b]	Methanol (metabolite)
Propane	Butane, isobutane, 2-propanol (metabolite[a])
2-Propanol	Acetone (metabolite)
1,1,1-Trichloroethane	Isopropyl nitrate (stabilizer[a])
2,2,2-Trichloroethanol	Trichloroethylene (also metabolite of chloral hydrate, dichloralphenazone and triclofos)
Trichloroethylene	2,2,2-Trichloroethanol (metabolite), chloroform [possibly from thermal degradation of trichloroacetic acid (metabolite) *in vitro*]
Xylenes	*ortho*-, *meta*- and *para*-Xylene occur together in technical xylene, *m*-xylene predominating. Ethylbenzene also contaminant in technical xylene

[a] Rarely found.
[b] Parent compound not normally detected in blood.
See Table 1 for full chemical names of certain compounds.

give rise to false positives. Processing blank frozen tissue can control for this possibility.

The interpretation of case data involving chloroform is particularly difficult, especially as this compound is still sometimes used in the course of crimes such as rape and murder. In addition to sometimes being present in drinking water at low concentrations, chloroform is found in a variety of medicinal preparations, in cigarette smoke, soft drinks, margarines, and in swimming pools if a chlorination plant is in operation. A further possible source of chloroform on headspace GC is from thermal decomposition of trichloroacetic acid. Trichloroacetic acid is a metabolite of several compounds, including the solvent trichloroethylene (Table 5) and the drugs chloral hydrate, dichloralphenazone and triclofos. Trichloroacetic acid has a half-life in blood of 3–5 days and thus may be detected for a relatively long time after exposure to, or ingestion of, a precursor. Trichloroacetic acid plasma concentrations of up to $40\,mg\,l^{-1}$ have been reported after occupational exposure to trichloroethylene vapor.

In 25 Caucasian adult women in Florida, USA, over a period of 6 months average plasma chloroform concentrations were generally less than $25\,\mu g\,l^{-1}$ but in two subjects plasma chloroform concentrations of 2.9 and $4.0\,mg\,l^{-1}$, respectively, were found during routine sampling. All subjects were carefully screened to exclude occupational and recreational exposure to chloroform and other compounds that could give rise to chloroform on headspace GC. At the other extreme, postmortem blood chloroform concentrations in fatalities involving this agent have been reported as $10–50\,mg\,l^{-1}$.

It is well known that ethanol may be both produced and metabolized by microbial action in biological specimens. Small amounts of hexanal may arise from degradation of fatty acids in blood on long-term storage, even at -5 to $-20°C$. Hexanal is resolved from toluene on the SPB-1 capillary GC system discussed above, but resolution may be lost if an isothermal quantitative analysis is performed. Interference from hexanal is only likely to be important, however, if very low concentrations of toluene ($0.1\,mg\,l^{-1}$ or less) are to be measured.

In some deaths attributed to the abuse of LPG, only butane, isobutane and propane are detected on headspace GC of postmortem samples. In other cases these three compounds are present, but in addition 2-propanol, acetone, 2-methyl-2-propanol, 2-butanol and/

or butanone are present (Fig. 3). By analogy with the metabolism of hexane (Table 5), these latter compounds probably arise from the metabolism of the butanes and propane (**Fig. 4**).

The alkyl nitrites which can be abused by inhalation (isobutyl nitrite, isopentyl nitrite) are a special case in that: (1) they are extremely unstable and break down rapidly *in vivo* to the corresponding alcohols (**Fig. 5**); and (2) usually also contain other isomers (butyl nitrite, pentyl nitrite). Any products submitted for analysis will usually contain the corresponding alcohols as well as the nitrites.

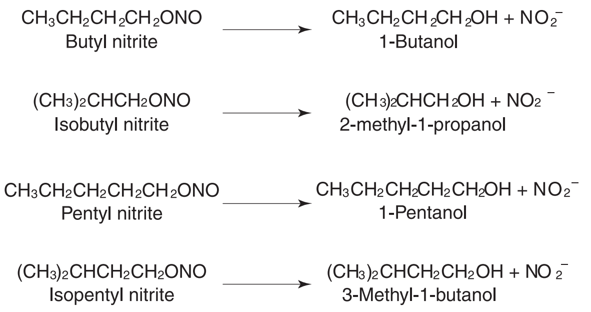


Figure 5 Breakdown of alkyl nitrites in humans. Note: pentyl = amyl.

Interpretation of quantitative results

Data to aid the interpretation of quantitative results in individual cases for a range of solvents, metabolites, etc. are given in **Table 7**. There may be a big overlap in the blood concentrations of volatile compounds attained after workplace exposure and as a result of deliberate inhalation of vapor. In the occupational setting, blood toluene concentrations after exposure to up to 127 p.p.m. toluene (UK occupational exposure limit at the time was 100 p.p.m.) for 8 h ranged between 0.4 and 6.7 $mg\,l^{-1}$. After brief exposure only signs of moderate intoxication (e.g. slurred speech, unsteady movements) have been associated with blood toluene concentrations as high as 30 $mg\,l^{-1}$. Blood toluene concentrations in samples from 132 patients who were thought to have engaged in VSA ranged from 0.2 to 70 $mg\,l^{-1}$, and were above 5 $mg\,l^{-1}$ in 22 of 25 deaths. On the other hand, 13 patients with blood toluene concentrations greater than 10 $mg\,l^{-1}$ were either asymptomatic or only mildly intoxicated (headache, nausea, vomiting and/or drowsiness), although these manifestations of toxicity can lead to 'indirect' acute VSA-related death, as discussed above. Aside from individual differences in tolerance and possible loss of toluene from the sample prior to analysis, sample contamination, etc., the lack of a strong correlation between blood concentrations and clinical features of poisoning is probably due to rapid initial tissue distribution and elimination.

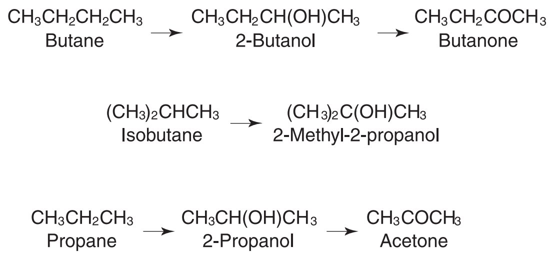


Figure 4 Summary of the metabolism of butane, isobutane and propane.

Some 80% of a dose of toluene is converted to hippuric acid (Table 5), which is excreted in urine. Similarly, more than 90% of a dose of xylene is metabolized to methylhippuric (toluric) acids. The principal isomer found in urine is 3-methylhippurate, as *m*-xylene is the principal component of technical grade xylene (Table 1). Methylhippurates are not normal urinary constituents, but hippuric acid may arise from the metabolism of benzoates in foods and medicines, and thus caution is needed in the interpretation of results. Hippurate and methylhippurate excretion is often expressed as a ratio to creatinine, as this obviates the need for 24 h urine collections. Occupational exposure to toluene can give rise to ratios of up to 1 g hippurate per gram creatinine, or more; in patients suspected of VSA a ratio of more than 1 g hippurate per gram creatinine strongly suggests, but does not prove, toluene exposure. Measurement of urinary *o*-cresol has been proposed as an alternative means of monitoring toluene exposure selectively, particularly in occupational circumstances, but the assay procedure is relatively complex and is thus not widely used.

After exposure to 350 p.p.m. 1,1,1-trichloroethane (UK maximum exposure limit at the time) for 1 h, the mean blood 1,1,1-trichloroethane concentration was 2.6 $mg\,l^{-1}$. Blood 1,1,1-trichloroethane concentrations ranged from 0.1 to 60 $mg\,l^{-1}$ in samples from 66 patients suspected of VSA, 29 of whom died. There was a broad relationship between blood 1,1,1-trichloroethane concentration and the severity of poisoning, but as in the case of toluene there was a big overlap between the blood concentrations encountered in fatalities and those attained after occupational exposure.

See also: **Analytical Techniques**: Mass Spectrometry; Gas Chromatography; Spectroscopic Techniques. **Drugs of Abuse**: Classification, including Commercial Drugs; Antemortem Blood; Postmortem Blood; **Pharmacology**.

Table 7 Solvents and other volatile compounds: interpretation of analytical toxicology results and mass/amount concentration conversion factors

Analyte	*'Therapeutic' or 'normal' whole blood or plasma concentration (less than)*	*Whole blood or plasma concentration associated with serious toxicity*[a]	*Relative atomic or formula mass*	*Mass/amount conversion*[b]	*Amount/mass conversion*
Acetaldehyde	0.2 mg l^{-1}	Not known	44.1	mg l^{-1} × 22.7 = μmol l^{-1}	μmol l^{-1} × 0.044 = mg l^{-1}
Acetone (see also 2-Propanol[c]) (urine)	10 mg l^{-1} (1 g l^{-1} in ketosis) 10 mg l^{-1} (2 1 g l^{-1} in ketosis) 80 mg l^{-1}[d]	Not known	58.1	g l^{-1} × 17.21 = mmol l^{-1}	mmol l^{-1} × 0.058 = g l^{-1}
Acetonitrile (see also Cyanide[c])	—	50 mg l^{-1}	41.1	mg l^{-1} × 24.3 = μmol l^{-1}	μmol l^{-1} × 0.041 = mg l^{-1}
Benzene	0.2 μg l^{-1} (nonsmokers) 0.6 μg l^{-1} (smokers)	1000 μg l^{-1}	78.1	μg l^{-1} × 12.8 = nmol l^{-1}	nmol l^{-1} × 0.078 = μg l^{-1}
Benzoate	0.01 g l^{-1} (dietary)	0.5 g l^{-1} (neonates)	122.1	g l^{-1} × 8.19 = mmol l^{-1}	mmol l^{-1} × 0.122 = g l^{-1}
Benzyl alcohol (see also Benzoate[c])	—	50 mg l^{-1} (neonates)	108.1	mg l^{-1} × 9.25 = μmol l^{-1}	μmol l^{-1} × 0.108 = mg l^{-1}
Bromide	10 mg l^{-1} (dietary)	500 mg l^{-1} (inorganic bromide exposure) 40 mg l^{-1} (organobromine alkylating agents)	79.9	mg l^{-1} × 12.52 = μmol l^{-1}	μmol l^{-1} × 0.080 = mg l^{-1}
Bromomethane: see Bromide[e]					
Butanone (urine)	10 mg l^{-1} (occupational exposure) 5 mg l^{-1}[d]	500 mg l^{-1}	72.1	mg l^{-1} × 13.9 = μmol l^{-1}	μmol l^{-1} × 0.072 = mg l^{-1}
Butoxyacetate (urine)	100 mg l^{-1}[d]		132.2	mg l^{-1} × 6.75 = μmol l^{-1}	μmol l^{-1} × 0.148 = mg l^{-1}
2-Butoxyethanol: see Butoxyacetate[e]					
Butyl nitrite: see Nitrite[c]					
Carbon disulfide: see 2-Thiothiazolidine-4-carboxylate[e]					
Carbon monoxide (carboxyhemoglobin saturation) (blood)	15% HbCO (heavy smokers)	20% HbCO	—		
Carbon tetrachloride	0.07 mg l^{-1}[d]	0.5 mg l^{-1} (2 h postexposure)	153.8	mg l^{-1} × 6.50 = μmol l^{-1}	μmol l^{-1} × 0.154 = mg l^{-1}
Chloroform	50 mg l^{-1} (anesthesia) 0.01 mg l^{-1} (dietary)	—	119.4	mg l^{-1} × 8.38 = μmol l^{-1}	μmol l^{-1} × 0.119 = mg l^{-1}
Cyanide (see also Thiocyanate[c])	0.2 mg l^{-1} (nonsmokers) 0.5 mg l^{-1} (heavy smokers)	5 mg l^{-1} (1 mg l^{-1} inhalation)	26.0	mg l^{-1} × 38.5 = μmol l^{-1}	μmol l^{-1} × 0.026 = mg l^{-1}
Cyclohexane	0.4 mg l^{-1} (occupational exposure)	Not known	84.2	mg l^{-1} × 11.9 = μmol l^{-1}	μmol l^{-1} × 0.084 = mg l^{-1}
Cyclopropane	100 mg l^{-1} (anesthesia)	—	42.1	mg l^{-1} × 23.8 = μmol l^{-1}	μmol l^{-1} × 0.042 = mg l^{-1}
Desflurane	20 mg l^{-1} (anesthesia)	—	168.0	mg l^{-1} × 5.95 = μmol l^{-1}	μmol l^{-1} × 0.168 = mg l^{-1}

Continued

Table 7 Continued

Analyte	*'Therapeutic' or 'normal' whole blood or plasma concentration (less than)*	*Whole blood or plasma concentration associated with serious toxicity*[a]	*Relative atomic or formula mass*	*Mass/amount conversion*[b]	*Amount/mass conversion*
1,1-Dichloro-1-fluoroethane	—	10 mg l^{-1}	117.0	mg l^{-1} × 8.55 = µmol l^{-1}	µmol l^{-1} × 0.117 = mg l^{-1}
Dichloromethane (see also Carbon monoxide[c])	1 mg l^{-1}[d]	200 mg l^{-1}	84.9	mg l^{-1} × 11.8 = µmol l^{-1}	µmol l^{-1} × 0.085 = mg l^{-1}
Diethyl ether	500 mg l^{-1} (anesthesia)	—	74.1	g l^{-1} × 13.50 = mmol l^{-1}	mmol l^{-1} × 0.074 = g l^{-1}
N,N-Dimethylformamide: see N-Methylformamide (NMF)[e]					
1,4-Dioxane	12 mg l^{-1} (occupational exposure)	Not known	88.1	mg l^{-1} × 11.4 = µmol l^{-1}	µmol l^{-1} × 0.088 = mg l^{-1}
β-Hydroxyethoxyacetate (urine)[e]	0.5 g l^{-1} (occupational exposure)		120.1	g l^{-1} × 8.33 = mmol l^{-1}	mmol l^{-1} × 0.120 = g l^{-1}
Enflurane	100 mg l^{-1} (anesthesia)	—	184.5	mg l^{-1} × 5.42 = µmol l^{-1}	µmol l^{-1} × 0.185 = mg l^{-1}
Ethanol	–	0.5 g l^{-1} (children) 2 g l^{-1} (adolescents/adults)	46.1	g l^{-1} × 21.7 = mmol l^{-1}	mmol l^{-1} × 0.046 = g l^{-1}
Ethoxyacetate (urine)	50 mg l^{-1}[d]		104.1	mg l^{-1} × 9.61 = µmol l^{-1}	µmol l^{-1} × 0.104 = mg l^{-1}
Ethoxyethanol: see Ethoxyacetate[e]					
2-Ethoxyethyl acetate: see Ethoxyacetate[e]					
Ethyl acetate: see Ethanol[c]					
Ethylbenzene: see Mandelate[e]					
Fluoride (serum) (urine)	0.2 mg l^{-1} 2 mg l^{-1} 4 mg l^{-1}[d]	2 mg l^{-1}	19.0	mg l^{-1} × 52.6 = µmol l^{-1}	µmol l^{-1} × 0.019 = mg l^{-1}
Formaldehyde (see also Formate[c])	4 mg l^{-1} (occupational exposure)		30.0	mg l^{-1} × 33.3 = µmol l^{-1}	µmol l^{-1} × 0.030 = mg l^{-1}
Formate (plasma)	10 mg l^{-1}		46.0	mg l^{-1} × 21.7 = µmol l^{-1}	µmol l^{-1} × 0.046 = mg l^{-1}
GHB: see 4-Hydroxybutyrate					
Glycol ethers: see 2-Butoxyethanol, 2-Ethoxyethanol					
Halothane	50 mg l^{-1} (anesthesia)	—	197.4	mg l^{-1} × 5.06 = µmol l^{-1}	µmol l^{-1} × 0.197 = mg l^{-1}
Trifluoroacetate (urine)[e]	2.5 mg l^{-1}[d]		114.0	mg l^{-1} × 8.77 = µmol l^{-1}	µmol l^{-1} × 0.114 = mg l^{-1}
Hexane (n-Hexane) (see also 2,5-Hexanedione[e])	10 µg l^{-1} (environmental)	Not known	86.2	µg l^{-1} × 11.6 = nmol l^{-1}	nmol l^{-1} × 0.086 = µg l^{-1}
2,5-Hexanedione (+ 4,5-Dihydroxy-2-hexanone) (urine)	5 mg l^{-1}[d]		[g]		
2-Hexanone: see 2,5-Hexanedione[e]					
Hippurate (urine)	0.2 g l^{-1} 2 g l^{-1}[h]		179.2	g l^{-1} × 5.59 = mmol l^{-1}	mmol l^{-1} × 0.179 = g l^{-1}
4-Hydroxybutyrate	—	250 mg l^{-1}	104.1	mg l^{-1} × 9.61 = µmol l^{-1}	µmol l^{-1} × 0.104 = mg l^{-1}
Isoamyl nitrite: see Nitrite[c]					

Table 7 Continued

Isobutyl nitrite: see Nitrite[c]					
MEK: see Butanone					
Isoflurane	20 mg l^{-1} (anesthesia)	—	184.5	mg l^{-1} × 5.42 = μmol l^{-1}	μmol l^{-1} × 0.185 = mg l^{-1}
Isopropanol: see 2-Propanol					
Mandelate (urine)	0.005 g l^{-1} 2 g l^{-1}[d]		152.1	g l^{-1} × 6.57 = mmol l^{-1}	mmol l^{-1} × 0.152 = g l^{-1}
MBK: see 2-Hexanone					
MEK: see Butanone					
Methanol (see also Formate[c])	0.002 g l^{-1} (dietary)	0.5 g l^{-1} (2 h postingestion) 0.2 g l^{-1} (6 h postingestion)	32.0	g l^{-1} × 31.25 = mmol l^{-1}	mmol l^{-1} × 0.032 = g l^{-1}
(urine)	0.03 g l^{-1}[d]				
Methoxyflurane	200 mg l^{-1} (anaesthesia)	—	165.0	mg l^{-1} × 6.06 = μmol l^{-1}	μmol l^{-1} × 0.165 = mg l^{-1}
N-Methylacetamide (NMA) (urine)	65 mg l^{-1}[f]		73.1	mg l^{-1} × 13.7 = μmol l^{-1}	μmol l^{-1} × 0.073 = mg l^{-1}
Methyl bromide: see Bromide[c]					
Methyl *n*-butyl ketone: see 2-Hexanone					
Methyl ethyl ketone: see Butanone					
N-Methylformamide (NMF) (urine)	15 mg l^{-1}[d]		59.1	mg l^{-1} × 16.9 = μmol l^{-1}	μmol l^{-1} × 0.059 = mg l^{-1}
Methylhippurates (total, urine)	2 g l^{-1}[d]		193.2	g l^{-1} × 5.18 = mmol l^{-1}	mmol l^{-1} × 0.193 = g l^{-1}
Methyl isobutyl ketone: see 4-Methyl-2-pentanone					
4-Methyl-2-pentanone (urine)	3.5 mg l^{-1}[d]		100.2	mg l^{-1} × 9.98 = μmol l^{-1}	μmol l^{-1} × 0.100 = mg l^{-1}
MIBK: see 4-Methyl-2-pentanone					
Nitrite	2.5 mg l^{-1}	Not known	46.0	mg l^{-1} × 21.7 = μmol l^{-1}	μmol l^{-1} × 0.046 = mg l^{-1}
(urine)	—	10 mg l^{-1}			
Nitrous oxide	100 mg l^{-1} (anesthesia)	—	44.0	mg l^{-1} × 22.7 = μmol l^{-1}	μmol l^{-1} × 0.044 = mg l^{-1}
NMA: see Dimethylacetamide					
NMF: see *N*, *N*-Dimethylformamide					
2-Propanol (see also Acetone[c])	—	2.5 g l^{-1}	60.1	g l^{-1} × 16.63 = mmol l^{-1}	mmol l^{-1} × 0.060 = g l^{-1}
Sevoflurane	50 mg l^{-1}	—	200.1	mg l^{-1} × 5.00 = μmol l^{-1}	μmol l^{-1} × 0.200 = mg l^{-1}
Tetrachloroethylene	1 mg l^{-1}[d]	10 mg l^{-1}	165.8	mg l^{-1} × 6.03 = μmol l^{-1}	μmol l^{-1} × 0.166 = mg l^{-1}
Tetrahydrofuran (urine)	8 mg l^{-1}[d]		72.1	mg l^{-1} × 13.9 = μmol l^{-1}	μmol l^{-1} × 0.072 = mg l^{-1}
Thiocyanate (plasma)	4 mg l^{-1} (nonsmokers) 20 mg l^{-1} (heavy smokers) 100 mg l^{-1} (nitroprusside therapy)	120 mg l^{-1}	58.1	mg l^{-1} × 17.2 = μmol l^{-1}	μmol l^{-1} × 0.058 = mg l^{-1}
2-Thiothiazolidine-4-carboxylate (TTCA) (urine)	8 mg l^{-1}[d]		163.2	mg l^{-1} × 6.13 = μmol l^{-1}	μmol l^{-1} × 0.163 = mg l^{-1}
Toluene (see also Hippurate[e])	1 mg l^{-1}[d]	10 mg l^{-1}	92.1	mg l^{-1} × 10.86 = μmol l^{-1}	μmol l^{-1} × 0.092 = mg l^{-1}
2-Methylphenol (urine)[e]	3 mg l^{-1}[d]				
2,2,2-Tribromoethanol	—	50 mg l^{-1}	282.8	mg l^{-1} × 3.54 = μmol l^{-1}	μmol l^{-1} × 0.283 = mg l^{-1}
Trichloroacetate (urine)	100 mg l^{-1}[d]		163.4	mg l^{-1} × 6.12 = μmol l^{-1}	μmol l^{-1} × 0.163 = mg l^{-1}
1,1,1-Trichloroethane	0.5 mg l^{-1}[d]	10 mg l^{-1}	133.4	mg l^{-1} × 7.50 = μmol l^{-1}	μmol l^{-1} × 0.133 = mg l^{-1}

Continued

Table 7 Continued

Analyte	*'Therapeutic' or 'normal' whole blood or plasma concentration (less than)*	*Whole blood or plasma concentration associated with serious toxicity*[a]	*Relative atomic or formula mass*	*Mass/amount conversion*[b]	*Amount/mass conversion*
2,2,2-Trichloroethanol	5 mg l^{-1} [d]	50 mg l^{-1}	149.4	mg l^{-1} × 6.69 = μmol l^{-1}	μmol l^{-1} × 0.149 = mg l^{-1}
Trichloroethylene	—	10 mg l^{-1}	131.4	mg l^{-1} × 7.61 = μmol l^{-1}	μmol l^{-1} × 0.131 = mg l^{-1}
(see also Trichloroacetate[e], 2,2,2-Trichloroethanol[c])					
TTCA: see 2-Thiothiazolidine-4-carboxylate					
Xylenes (total) (see also Methylhippurates[e])	0.01 mg l^{-1} (environmental) 1.5 mg l^{-1} [d]	Not known	106.2	mg l^{-1} × 9.42 = μmol l^{-1}	μmol l^{-1} × 0.106 = mg l^{-1}

[a] Dash indicates that concentration which could be associated with serious toxicity may be no greater than 'therapeutic' concentration in other patients.
[b] The equation a mg l^{-1} × (conversion factor) = *b* μmol l^{-1}, for example, and vice versa, is implied.
[c] Active metabolite/decomposition product.
[d] Deutsche Forschungsgemeinschaft suggested 'action level' 1996.
[e] Metabolite/decomposition product.
[f] Suggested UK Health and Safety Executive 'health guidance value' 1996.
[g] No accurate relative molecular mass.
[h] Suggestive of recent toluene exposure, but dietary benzoate may give elevated hippurate excretion.

Further Reading

de Zeeuw RA, Franke JP, Machata G *et al.* (1992) *Gas Chromatographic Retention Indices of Solvents and Other Volatile Substances for Use in Toxicological Analysis*. Report XIX of the DFG Commission for Clinical-Toxicological Analysis. Weinheim: VCH.

Flanagan RJ and Ives RJ (1994) Volatile substance abuse. *Bulletin on Narcotics* **46**:49–78.

Flanagan RJ and Meredith TJ (1996) Volatile substances with special reference to volatile substance abuse. In: Descotes J (ed.) *Human Toxicology*, pp. 577–622. Amsterdam: Elsevier.

Flanagan RJ, Streete PJ and Ramsey JD (1997) *Volatile Substance Abuse*. United Nations International Drug Control Programme Technical Series Number 5. Vienna: UNIDCP.

Gill R, Hatchett SE, Osselton MD, Wilson HK and Ramsey JD (1988) Sample handling and storage for the quantitative analysis of volatile compounds in blood: the determination of toluene by headspace gas chromatography. *Journal of Analytical Toxicology*, **12**: pp. 141–146.

Gill R, Hatchett SE, Warner HE *et al.* (1992) The quantitative analysis of volatile solvents and gases in blood samples: problems and solutions. *Proceedings of the 26th International Meeting of the International Association of Forensic Toxicologists, Glasgow, 1989*, pp. 372–382. Edinburgh: Scottish Academic Press.

Health and Safety Executive (1998) *EH40/98 Occupational Exposure Limits 1998*. Norwich: HMSO.

Kozel N, Sloboda Z, de la Rosa M (eds) (1995) *Epidemiology of Inhalant Abuse: An International Perspective*. NIDA Research Monograph 148. Rockville: National Institute on Drug Abuse.

Ojanperä I, Pihlainen K and Vuori E (1998) Identification limits for volatile organic compounds in the blood by purge-and-trap GC-FTIR. *Journal of Analytical Toxicology* **22**:290–295.

Osterloh J (1984) Butyl nitrite: analytical techniques and toxicology. In: Baselt RC (ed.) *Advances in Analytical Toxicology*, vol 1, pp. 159–197. Davis, CA: Biomedical Publications.

Payne JP (1998) The criminal use of chloroform. *Anaesthesia* **53**:685–690.

Pihlainen K and Ojanperä I (1998) Analytical toxicology of fluorinated inhalation anaesthetics. *Forensic Science International* **97**:117–133.

Streete PJ, Ruprah M, Ramsey JD and Flanagan RJ (1992) Detection and identification of volatile substances by head-space capillary gas chromatography to aid the diagnosis of acute poisoning. *Analyst* **117**:1111–1127.

Taylor JC, Norman CL, Bland JM, Ramsey JD and Anderson HR (1998) *Trends in Deaths Associated with Abuse of Volatile Substances 1972–1996*. London: St George's Hospital Medical School.

Uyanik U (1997) Gas chromatography in anaesthesia I. A brief review of analytical methods and gas chromatographic detector and column systems. *Journal of Chromatography* **693**:1–9.

Winek CL, Wahba WW and Rozin L (1995) Accidental death by nitrous oxide inhalation. *Forensic Science International*, 73: pp. 139–141.

Interpretation of Results

W Lee Hearn, Miami-Dade County Medical Examiner Department, Miami, FL, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0546

Introduction

Forensic toxicology concerns the analysis of biological specimens (fluids and tissues) for the presence and, often, the concentration of drugs and poisons. The results of the analyses must be correlated with the circumstances of the case to determine what role, if any, the detected substances played. This correlative function is commonly called interpretation. This article will examine the three major subspecialties of forensic toxicology and the various factors that enter into the interpretation process in each.

The forensic toxicology laboratory exists for the sole purpose of providing interpretable analytical data. Therefore, the analytical strategy is designed with anticipation of the need for later interpretation. The most appropriate specimens should be analyzed by sensitive, specific and quantitatively accurate and precise techniques to yield reliable data upon which to base opinions. The toxicologist must be absolutely certain that the analytical data are accurate. Furthermore, the samples must be properly preserved and clearly traceable to the subject of the investigation by an unbroken chain of custody, and handled and stored with a level of security sufficient to preclude tampering.

The choice of specimen(s) and the scope of analysis are determined largely by the purpose of the investigation. Modern forensic toxicology can be divided into three major categories: forensic urine drug testing (FUDT), human performance toxicology and postmortem toxicology. FUDT seeks evidence of illegal drug use by current or prospective employees; human performance toxicology attempts to determine whether the subject was impaired or intoxicated at some specific time by analyzing specimen(s) collected later; and postmortem toxicology investigates the role of drugs and poisons in causing or contributing to the subject's death. Each presents specific requirements and challenges for the interpreting toxicologist.

Forensic Urine Drug Testing

FUDT, also known as workplace drug testing, has as its goal the minimization of drug abuse in the workplace. This goal is accomplished by intimidating potential drug users through fear of detection and by elimination of drug users from the workforce through treatment or discharge. Since most employees are not suspected of any wrongdoing, the sampling process is designed to be minimally intrusive. Urine is usually the specimen of choice, although hair and sweat are undergoing evaluation as potential alternative specimens. To control the cost of analysis, FUDT programs restrict the scope of analysis to the drugs deemed most dangerous by virtue of their addictive nature, illegality, abuse liability or potential harm to the employees' health or productivity. The most commonly included drugs are methamphetamine, amphetamine, cocaine, marijuana, opiate narcotics (morphine, codeine, heroin) and phencyclidine (PCP). Other drugs such as lysergic acid diethylamide (LSD) and methylenedioxymethamphetamine (MDMA; 'ecstasy') may be added if their use is perceived to be common.

In workplace drug testing, the only issue is whether or not the subject of the test illegally used a controlled drug. The drug testing is usually done on urine. If a drug or drug metabolite(s) is conclusively identified, in a properly collected sample, then the person who provided the urine must have ingested the drug. However, legitimate questions may be raised about whether ingestion was intentional or unknowing. Such questions are often referred to a physician trained in interpretation of drug testing reports, a medical review officer (MRO), but they may also be asked of the toxicologist. The toxicologist or the MRO may be able to comment on the reasonableness of arguments offered in defense of one who tested positive for a drug. Above all, unbiased interpretation must include consideration of alternative explanations for apparently incriminating findings.

First, the toxicologist or MRO must be aware of circumstances that can produce positive tests in a person who has not abused the drug (**Table 1**). For example, research has shown that passive exposure to marijuana smoke does not yield urinary concentrations of Δ^9-tetrahydrocannabinol (THC) metabolites above $40\,ng\,ml^{-1}$. However, taking dietary supplements containing hemp seed oil or eating hemp seed

Table 1 Innocent causes of positive urine drug tests.

Drug	*Source*
Marijuana	Hemp seed oil or hemp seed confections
Cocaine	Coca tea, maté de coca, may be imported from South America and has been available in health food stores in the past
Morphine, heroin, codeine	Poppy seed pastry – contains morphine and codeine
Amphetamine	Some phenylpropanolamine products as a byproduct of manufacture Metabolism of medications, e.g. clobenzorex
Methamphetamine	Nasal inhalers: *l*-deoxyephedrine= methamphetamine; metabolism of medications, e.g. deprenyl=selegiline

confections can cause readings over $200\,\mathrm{ng\,ml^{-1}}$. Over-the-counter diet products containing phenylpropanolamine may have traces of amphetamine, produced as a byproduct of manufacture. If a trace of amphetamine is found with a high concentration of phenylpropanolamine, ingestion of the amphetamine may have been inadvertent. Poppy seeds contain traces of morphine sufficient to cause a positive drug test if poppy seed pastries are eaten. Quite high urine morphine concentrations have been obtained in controlled experiments in which human volunteers provided urine samples for opiate analysis after consuming poppy seed pastries. Testing for the alkaloid, thebaine, derived from poppy seeds, may help to distinguish food from drug. Several pharmaceuticals contain, or are metabolized to, the 'l' isomer of methamphetamine (i.e. l-deoxyephedrine), rather than the controlled 'd' isomer. Routinely applied confirmatory tests can not distinguish between optical isomers, but chiral separation columns or derivatives can. Such definitive testing may be required if someone challenges a positive urine test for methamphetamine. Another commonly reported explanation for positive urine drug test results is the subject's unknowing consumption of a food or beverage containing a drug that someone else added surreptitiously (e.g. marijuana brownies). The testing methods are sufficiently sensitive to detect excretion products from doses of drug that may be too small to produce observable symptoms. The possible motivation for someone to administer the drug should be considered, as well as the opportunity to do so without detection. Women sometimes allege that they were exposed to a drug such as cocaine through transfer of semen from a drug-using sex partner; however, the amount of cocaine in semen of a user has been measured and is insufficient to produce a positive urine test if ingested. Some people have suggested that their cocaine-positive drug test resulted from handling contaminated money. The amount of cocaine found in US currency is too small to cause a positive test by dermal absorption, but a single cocaine particle falling into the sample container from the subject's hand could produce a positive result. To prevent contamination, test subjects are often required to wash their hands before providing a urine sample.

Human Performance Toxicology

Human performance toxicology deals with the effects of alcohol and other intoxicating substances on a subject's ability to drive a motor vehicle or engage in other potentially hazardous activities. When impairment, rather than use, is the issue, more information is needed for interpretation. Presence of drug or metabolites in urine is useless for demonstrating impairment. At best, finding an intoxicating substance in urine can explain the source of obvious symptoms of impairment. The concentration of drug and its active metabolites(s) in blood is important, but other factors must also be considered. Individual variability and tolerance can affect the degree of impairment at any given blood drug concentration. Combinations of drugs can interact to increase or decrease impairment. For most drugs, a relationship between drug concentration and performance impairment has not been established. Therefore, evidence for interpretation should include the subject's blood concentrations of all drugs capable of causing impairment, as well as evidence that the subject was impaired. The toxicologist familiar with the behavioral effects of drugs may be able only to offer an opinion that the observed symptoms are consistent, or not consistent, with the effects of the detected drugs. However, to infer that because a drug was present, regardless of concentration, the subject must have been impaired without having evidence of symptoms of intoxication is speculation, not objective interpretation.

Drug recognition programs

Drug recognition (DR) programs, such as that sponsored by the US National Highway Traffic Safety Administration, can provide valuable documentation of intoxication. In DR programs, trained observers examine subjects suspected of being intoxicated, and

record their observations in a standardized format. The examination includes physiological symptoms (e.g. pulse, respiration, pupil responses and others), physical tests of coordination and balance, and tests of mental agility, memory and perception. Finally, an algorithmic approach is applied to the observations to place symptom patterns into eight categories: not impaired, or impaired by sedative, stimulant, narcotic, hallucinogen, marijuana, PCP-like or volatile inhalant substances. If the subject is examined while symptoms are present, the documentary evidence can be evaluated in conjunction with results of toxicological analyses. The forensic toxicologist may then be able to offer opinions on the cause-and-effect relationship of the observed behavior or symptoms and the detected intoxicants. Research has shown accuracy rates in the range of 50–80% for DR examinations, depending upon the drug category and the examiner's training and experience. Therefore, DR reports must always be confirmed by laboratory analysis before any punitive action, such as conviction for driving while intoxicated, can be justified.

Postmortem Toxicology

Because of their variety and complexity, postmortem toxicological investigations provide the greatest challenges for the interpreting toxicologist. The questions that may be raised in a death investigation may be the same as those involved in workplace drug testing or human performance toxicology, but often the postmortem toxicologist must deal with far more complicated issues. To do so requires a familiarity with principles of pharmacology, physiology, biochemistry and anatomy, but also an awareness of pathological changes associated with toxicity and postmortem changes affecting drug concentrations. Knowledge of factors such as the behavioral effects of drugs and the lifestyle of drug-using subcultures can help in understanding and explaining some aspects of a case.

In postmortem toxicology, interpretation is a thought process that begins with the initial case review and guides the analytical strategy that culminates with the toxicologist forming opinions regarding the involvement of drugs or other substances in the death of an individual. The first question that must be addressed is what, if anything, was ingested by the deceased? The answer lies in the outcome of a thorough toxicological analysis that is capable of detecting relevant (i.e. therapeutic or subtherapeutic) concentrations of any substance that could realistically be present. The extent of the search is governed, in part, by the other aspects of the investigation. If there is a strong suspicion of a drug or poison then the toxicologist must search more diligently than if there is no such suspicion. A negative result must be interpreted from an understanding of the sensitivity of the analytical method. It implies that the analyte targeted by the test was not present, or if present, was at a concentration below the detection limit. If methods of sufficient sensitivity are used, negative findings mean that either nothing was ingested or something outside the scope of analysis was involved. Knowledge of the chemical properties of various drugs and poisons and of the capabilities and limitations of available analytical techniques are essential to the determination of whether the analysis is capable of detecting all targeted substances.

When drugs or other toxic substances are detected in body fluids or tissues, the toxicologist must gather additional data in an effort to determine what role, if any, the detected substances played in the events leading to death. In some cases where the cause of death is manifestly obvious, such as accidental, the result of assault or self-inflicted trauma, the issue becomes one of the impairing or intoxicating effects of drugs, including alcohol (behavioral toxicity). Substances lacking psychoactivity are usually incidental to the investigation, unless they could have affected survivability. In cases of death attributable to natural disease, the toxicologist is concerned with substances that can exacerbate pre-existing illness. Examples include cocaine and other sympathomimetic agents with heart disease or berry aneurysms, and nonselective β-adrenergic blocking agents in patients with asthma. Other potential issues for the toxicologist include adverse interactions of therapeutic drugs and unsuspected intentional overdose in an otherwise seriously ill person. In some cases, such as death from epileptic seizures, it may be important to document whether the deceased was compliant with life-sustaining therapy.

When a toxic mechanism of death is suspected from the findings of the investigation and autopsy, the toxicologist is asked to identify the offending substance(s) and, insofar as possible, to determine the route and mode of entry. Was ingestion accidental or intentional? Was enough drug or poison present to cause death, and was it ingested acutely or cumulatively? If multiple substances are detected, do they possess additive or synergistic toxic activity?

Drug interactions

If multiple substances are identified, some may be unrelated to the cause of death, whereas others may have interacted, resulting in fatal consequences. Drugs with similar toxic mechanisms, such as respiratory depression or cardiovascular stimulation or depression, can be expected to contribute additively

or synergistically to the net toxic effect. For example, ethanol, benzodiazepines and opiates all have respiratory depressant activities at toxic concentrations. In combination they can produce serious or fatal respiratory depression, even when the individual drugs are present in subtoxic concentrations. The calcium channel blocker, verapamil, and β-adrenergic antagonists, such as propranolol or timolol, may both be prescribed for migraine headaches or to lower blood pressure. When combined, verapamil and beta blockers can sometimes lead to fatal cardiogenic shock and can even cause the heart to stop beating. It is sometimes apparent that one drug exerted the predominant effect, without which the death would not have occurred. This is often the case with heroin toxicity. Other drugs, such as alcohol, benzodiazepines, cocaine or other opiates, if present, would be expected to contribute to toxicity, but usually would not be lethal without the addition of heroin.

Postmortem changes affecting toxicology

During life, drugs are differentially concentrated in the various organs to levels that may be orders of magnitude higher than that in blood. In general, the higher the volume of blood into which a dose of the drug appears to distribute (volume of distribution), the greater is the difference in concentrations between blood and tissues. The liver, heart and lungs are among the tissues known to sequester various drugs. After death, the physiological and chemical conditions that maintain the differential begin to degrade and the drugs diffuse out of tissues into blood. Therefore, blood collected from the vena cava is likely to contain excess drug released by the liver, as well as any drug freshly absorbed from the gastrointestinal tract. Heart blood, 'chest cavity blood' or aorta blood may contain drug released by the heart tissue, lungs and liver and, potentially, may be contaminated by stomach contents. A therapeutic concentration of drug can quickly rise into the 'toxic' or 'lethal' range on tables derived from clinical observations of living subjects. Large differences in drug concentrations are often found when blood from these visceral sites is compared with blood from a peripheral site, such as femoral or subclavian vein (**Table 2**). It is now recognized that analysis of peripheral blood yields more consistent and more reliable quantitative data. However, the toxicologist can not simply compare a drug's concentration in blood with published clinical data of therapeutic, toxic and lethal concentrations. To do so would probably lead to erroneous conclusions. Postmortem drug concentrations should be compared with postmortem data on the same site from cases of known drug toxicity and from cases where toxicity was not involved. Such an approach will often help to distinguish elevated from normal drug levels. Analysis of tissue such as liver frequently reveals extremely high concentrations of many drugs in overdose cases. Likewise, measuring the total amount of unabsorbed drug remaining in the stomach after death can support the conclusion of toxicity if a quantity representing more than a normal dose is found. Finding a trace of drug in the stomach does not prove an oral route of administration, because most drugs diffuse from the blood into the stomach. The same is true when cocaine is detected in a sample obtained by swabbing the nasal passages. A small amount of cocaine will pass from blood to the nasal secretions regardless of the route of administration. Only if a relatively large amount of cocaine is found can one conclude that it was taken intranasally.

Table 2 Drugs implicated in postmortem redistribution.

Alprazolam	Imipramine
Amantadine	Lidocaine
Amitriptyline	Maprotiline
Amobarbital	Meperidine
Amoxapine	Mesoridazine
Amphetamine	Methadone
Antipyrine	Methamphetamine
Brompheniramine	Metoprolol
Butalbital	Nordiazepam
Caffeine	Nortriptyline
Chlordiazepoxide	Pentobarbital
Chlorpheniramine	Phencyclidine
Chlorpromazine	Phenobarbital
Cocaine	Phentermine
Codeine	Promethazine
Desipramine	Propoxyphene
Dextromethorphan	Secobarbital
Diazepam	Temazepam
Digoxin	Thioridazine
Diltiazem	Trazodone
Diphenhydramine	Triazolam
Doxepin	Trichloroethanol
Doxylamine	Trifluoroperazine
Ethclorvynol	Trimipramine
Fluoxetine	

While postmortem redistribution may elevate drug concentrations in blood, some drugs continue to be metabolized after death. Enzymes, especially esterases, that do not require oxygen or consume energy, remain active and hydrolyze susceptible drugs, including cocaine, heroin and a heroin metabolite, 6-monoacetylmorphine. The net effect of the competing processes, redistribution and metabolism, on drug concentrations in postmortem specimens is unpredictable.

Ethyl alcohol can be produced by fermentative microorganisms during putrefaction so that ethanol

concentrations in specimens from a decomposing body are not representative of the blood alcohol concentration at the time of death. In many cases where ethanol is found in a decomposed body, it may be said that the deceased probably consumed some alcohol before death, but when ethanol concentrations are lower than $0.1\,g\,dl^{-1}$ in a severely putrefied body, the alcohol may have been entirely produced by postmortem fermentation. Even higher values of postmortem ethanol production (over $0.2\,g\,dl^{-1}$) have been reported.

Putrefaction can also affect interpretation of carbon monoxide measurements in postmortem blood. The percent of total hemoglobin combined with carbon monoxide (i.e. carboxyhemoglobin) is usually measured spectrophotometrically. A decomposition product, sulfmethemoglobin, has an absorption spectrum that overlaps that of carboxyhemoglobin. Putrefaction will elevate the carbon monoxide reading with most methods, but may depress it with others. Therefore, the toxicologist must understand the effects of sulfmethemoglobin on the method whose result is being interpreted to know whether the reading is artifactually high or low in a decomposed sample.

Correlative information

The presumption of toxicity must be further evaluated, taking other physical and factual evidence into consideration. The terminal events, medical and social history, and autopsy findings must be correlated with drug concentration data. Tolerance from chronic use of a drug such as morphine can raise the threshold for therapeutic effect into a range that would be lethal to a nontolerant individual. Pulmonary edema and congestion revealed by autopsy can represent respiratory depression and lend support to a suspicion of drug toxicity. Even though many pathological changes in the body are not specifically associated with drug toxicity, their presence is supportive, whereas their absence would indicate that drug toxicity was not responsible for the death. Some diagnoses, such as sudden infant death syndrome (SIDS), and some cardiac deaths are justified only after other causes, including poisoning, are excluded.

Observations from the death scene, such as a body posture that restricts respiration or a plastic bag over the head, can explain why some people die with drug concentrations that, while elevated, would not normally be expected to cause death. Drug intoxication in these cases would be one factor that prevents the victim from restoring normal respiratory exchange. Elevated body temperature measured at a death scene can be related to a disturbance in thermoregulatory homeostasis that is often associated with toxicity due to cocaine, amphetamines, neuroleptics and some other drugs.

Witnesses to a death can sometimes provide valuable information about the behavior of the deceased and the terminal symptoms. Reports of bizarre and violent behavior followed by sudden death are evidence of excited delirium, which may be attributable to cocaine or amphetamines if supported by analytical documentation of their presence. Concentrations are usually below the range considered to be lethal, so the behavioral evidence may be critical in arriving at a correct conclusion. Drug paraphernalia found at a death scene raises or supports a suspicion of drug toxicity, which can be confirmed by autopsy and toxicological analysis.

When a hospitalized victim of trauma or poisoning dies in spite of treatment, medical records documenting the progression of symptoms can sometimes be correlated with postmortem toxicology data. The mechanism of toxicity of the drugs or poison, such as the electrocardiographic abnormalities caused by tricyclic antidepressants, should be manifest in the reported symptoms and observed physical findings. This becomes especially important when life support efforts delay the terminal event, allowing the responsible toxin to be eliminated to a point that postmortem concentrations are too low to be detected, or, if detected, to be considered toxic without the additional supporting evidence. In some jurisdictions, the postmortem toxicology laboratory obtains specimens of blood, urine and stomach contents collected soon after the patient's arrival at the hospital. These antemortem specimens sometimes yield the only reliable evidence of the deceased's state of intoxication at the time of admission. For example, drugs administered by physicians, including morphine, haloperidol, lignocaine (lidocaine), atropine and others, will not usually be found in admission specimens, whereas drugs taken prior to hospitalization will.

Drug metabolism (biotransformation)

Whenever a foreign substance is introduced into the body, drug-metabolizing enzymes act upon it to accelerate its elimination. Cytochromes of the P_{450}-class, esterases and other hydrolases introduce polar functional groups, such as hydroxyl, carboxyl or amino, and conjugating enzymes, including glucuronyl transferases and sulfotransferases, link the polar functional groups to highly water-soluble polar molecules, glucuronic acid or sulfate, in order to form water-soluble conjugates that are readily excreted.

Biotransformation of drugs and poisons is a factor that is frequently considered in the interpretation

process. Drug metabolites are often pharmacologically inactive, but some retain activity or possess activity different from the parent drug. In some cases such as aromatic amines (2-naphthyl amine, 4-aminobiphenyl), halogenated hydrocarbons (chloroform, carbon tetrachloride) and acetaminophen, metabolism yields toxic products. Some biotransformation reactions are subject to induction or inhibition by certain drugs so that the rate of metabolism of other drugs or poisons is increased or decreased. Thus, if an inhibitor of a specific enzyme is present, concentrations of that enzyme's substrate can rise to toxic levels, or an inducer can cause accumulation of a toxic metabolite or can lower a drug's concentration to an ineffective level. An awareness of metabolic interactions and biotransformations can help the toxicologist distinguish between intentional and accidental intoxication.

The postmortem toxicology laboratory often quantifies both parent drug and its principal metabolite(s). The ratio of parent drug to metabolite concentrations can indicate whether ingestion was chronic or acute. Metabolites of many drugs accumulate in the body with chronic administration. Conversely, an acute overdose may yield only a small amount of metabolite before death occurs. Furthermore, drugs such as propoxyphene and meperidine that are converted to toxic metabolites may produce toxicity by virtue of accumulation of the metabolite during chronic administration.

Pharmacokinetics

The science of pharmacokinetics describes the time course of drug action in terms of the processes of absorption, distribution and elimination. The onset of drug action is determined, first, by the rate of distribution to the site of action. The action is terminated by the elimination of the drug via metabolism and excretion.

Although the concentration of drug in the postmortem blood, tissues or urine cannot usually be related to pharmacokinetic parameters determined in living humans, the clinical data are still valuable to the postmortem toxicologist. For example, a drug's elimination rate can usually be expressed in terms of its half-life ($t_{\frac{1}{2}}$ elim.), which is the length of time required for the concentration to decline to half of its initial value in a process following first-order kinetics (i.e. the rate of elimination is proportional to drug concentration at any time point). Approximately 99% of the drug will be eliminated by 6–7 half-lives after its distribution is complete. Usually absorption and distribution occur much faster than elimination. Therefore, if a drug is detected in postmortem blood, it was probably taken within 6–7 half-lives prior to death.

The distribution pattern between blood and other biofluids and tissues can also be useful for interpretation. Probably the drug whose pharmacokinetic behavior is best understood is ethanol. In postmortem toxicology laboratories, ethanol is usually measured in different body fluids from the same individual to facilitate interpretation. At or near equilibrium, the concentration of ethanol in vitreous humor (eye fluid) is usually about 15–20% higher than that in whole blood. If both fluids are analyzed and blood has higher ethanol concentration than vitreous, then by the pharmacokinetics of ethanol (**Fig. 1**), the individual was in the absorption phase, or the sample was contaminated with ethanol from stomach contents or putrefaction. Conversely, if the vitreous ethanol concentration is much higher than the blood, the blood sample may have been diluted, and the toxicologist should consider the site of collection and whether intravenous fluids were administered prior to death. To resolve discrepancies, other samples may be analyzed, including blood from another site, bile, brain tissue, urine and stomach contents. A pattern should emerge from the data to explain the pharmacokinetic state at the time of death.

Behavioral toxicity

Psychoactive drugs, both licit and illicit, can cause alterations in perception, judgment, coordination, mood and information processing. In cases involving violence or accidental trauma, when such drugs are detected in the body of the deceased, questions inevitably arise regarding the role of the drug(s) in

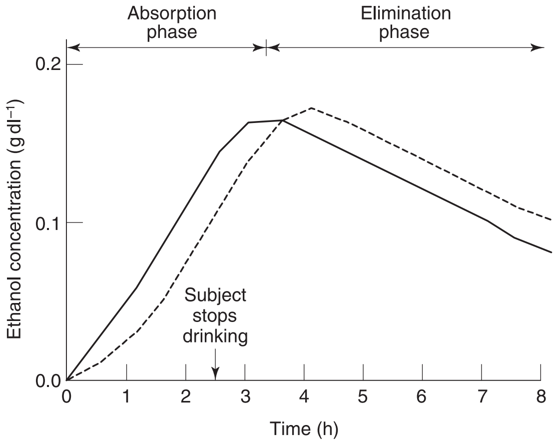


Figure 1 Hypothetical pharmacokinetic curves for ethyl alcohol in blood (——) and vitreous humor (----). Note that the vitreous alcohol concentration is lower than that in the blood during the absorption phase, but it exceeds the blood alcohol concentration during elimination.

precipitating the fatal event. The issues are similar to those in human performance toxicology, except that specimens are usually collected postmortem rather than from a living subject. Therefore, drug concentrations must be considered in light of postmortem changes, and impairment or intoxication must be deduced from factual evidence (eyewitnesses) and physical evidence available at the scene where the fatal incident occurred. For example, a motor vehicle crash investigation may discover evidence of driver error, and later toxicological analysis reveals elevated blood alcohol or a high concentration of a drug with the potential to cause impairment. Similarly, when a person becomes excited, violent and irrational, then disrobes, breaks objects, struggles with police and suddenly becomes unresponsive and dies, a finding of even a few hundred nanograms per milliliter of cocaine in the deceased's blood suggests that the death can be attributed to cocaine-induced excited delirium. This opinion is supported by a documented high postmortem body temperature and a postmortem blood benzoylecgonine concentration of several thousand nanograms per millimeter.

Because of the many factors affecting measured postmortem drug concentrations, such data can rarely serve as the sole basis for interpretation. They must be correlated with evidence from the autopsy and with information gathered by police, and other investigators. Conversely, inferences and suspicions derived from autopsy and investigation must be refined and confirmed thorough toxicological studies. Only then can reliable opinions be formulated regarding the role of drugs or poisons in a death investigation.

See also: **Alcohol**: Blood; Body Fluids; Interpretation. **Drugs of Abuse**: Antemortem Blood; Postmortem Blood; Urine. **Toxicology**: Overview; Methods of Analysis – Ante Mortem; Methods of Analysis – Post Mortem. **Pharmacology**.

Further Reading

Baselt RC and Cravey RH (1995) *Disposition of Toxic Drugs and Chemicals in Man*, 4th edn. Foster City, CA: Chemical Toxicology Institute.

Druid H and Holmgren P (1997) A compilation of fatal and control concentrations of drugs in postmortem femoral blood. *Journal of Forensic Sciences* **42**:79–87.

Ellenhorn MJ, Schonwald S, Ordog G and Wasserberger J (1997) *Ellenhorn's Medical Toxicology: Diagnosis and Treatment of Human Poisoning*, 2nd edn. Baltimore, MD: Williams and Wilkins.

Garriott JC (ed.) (1993) *Medicolegal Aspects of Alcohol Determination in Biological Specimens*. Phoenix, AZ: Lawyers and Judges Publishing.

Green KB and Isenschmid DS (1995) Medical review officer interpretation of urine drug test results. *Forensic Science Reviews* 7:41–60.

Karch SB (ed.) (1998) *Drug Abuse Handbook*. Boca Raton, FL: CRC Press.

Klaassen CD (1996) *Casarett and Doull's Toxicology: The Basic Science of Poisons*, 5th edn. New York: McGraw-Hill.

Liu RH (1992) Important considerations in the interpretation of forensic urine drug test results. *Forensic Science Reviews* **4**:52–64.

Liu RH and Goldberger BA (eds) (1995) *Handbook of Workplace Drug Testing*. Washington, DC: AACC Press.

O'Hanlon JF and de Gier JJ (1986) *Drugs and Driving*. London: Taylor and Francis.

Olson KR (ed.) (1996) *Poisoning and Drug Overdose*. Norwalk, CT: Appleton and Lange.

Sunshine I (ed.) (1989–1993) *Yearbook of Toxicology* (series). Boca Raton, FL: CRC Press.

Methods of Analysis – Ante Mortem

O H Drummer, Victoria Institute of Forensic Medicine and Monash University, Melbourne, Australia

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0544

Introduction

Drug screening in antemortem specimens of forensic cases has many uses, all of which assist the investigating authorities in providing relevant information pertaining to the cases (**Table 1**). Ultimately, toxicology testing results will assist the courts in establishing the truth by either providing evidence of drug use, or by refuting the use of relevant drugs. This latter observation is important, as drug use is often suspected, but can only be confirmed by toxicology testing procedures.

Toxicology testing is particularly important in victims of sexual assault, who may have been given drugs by the assailant to reduce consciousness and memory. Drugs used in these cases are typically one of the benzodiazepines (flunitrazepam, alprazolam, etc.), barbiturates and, more recently, γ-hydroxybutyrate (GHB). Perpetrators of violent crime may

Table 1 Reasons for drug testing in forensic cases.

Establishing drug use in victims of sexual and physical assaults
Establishing drug use in drivers of motor vehicles
Establishing drug use in persons involved in workplace accidents
Establishing workplace or environmental exposure of workers
Assisting investigations for deaths occurring in hospital
Assisting investigators with estimation of timing of drug use

also have consumed alcohol or illicit drugs, or may even be under medication. In practice, drug users committing crimes are likely to be under the influence of two or more drugs. Drivers involved in motor vehicle accidents or traffic infringements are also frequently under the influence of two or more drugs. Toxicology testing on specimens taken soon after the incident may assist in establishing the sobriety, or otherwise, of these drivers at the time of the accident or traffic infringement. Drug testing in antemortem specimens of persons dying in hospital, and who come under a medical examination order, reduces the problems associated with postmortem redistribution. Postmortem processes can falsely elevate blood concentrations, frustrating any interpretation of postmortem toxicology.

Since the great majority of cases ($>70\%$) involve more than one drug, it is advisable to conduct a broad drug screen to include most of the common drugs of abuse, rather than target the analysis to one or a limited range of drugs suggested by the circumstances.

Specimens

Specimens collected ante mortem are most often blood, for plasma or serum, or urine; however, other specimens, such as hair, sweat, and saliva, have also been used to assess drug use.

Blood and plasma

Blood contains predominately red blood cells, white cells, and plasma. Plasma is obtained from non clotted blood by removal of the cells following centrifugation; serum is the liquid phase remaining after blood is allowed to clot. In this article, plasma and serum are considered to be one specimen, unless otherwise differentiated.

Blood, or plasma, or serum derived from blood, is the most useful specimen that can be collected, as drugs present in this fluid can best be related to a physiological effect and can be used to assess the likelihood of recent drug use or exposure to chemicals. Programs of therapeutic drug-monitoring of plasma are frequently conducted in clinical toxicological laboratories and form the basis of therapeutic drug compliance and optimization of drug doses. Typically, immunoassays are used in drug monitoring and screening, although high-performance liquid chromatography (HPLC), gas chromatography (GC), and mass spectrometry (MS) techniques are equally well suited.

Urine

This is, with blood or plasma, the most frequently collected specimen. Since concentrations of drugs and metabolites of drugs are usually much higher than in blood, urine provides a valuable specimen for assessing drug use over the previous day or two. Relatively large volumes (50 ml or more) can be collected, allowing sufficient specimen even for less sensitive techniques; however, the presence of a drug in urine does not necessarily imply recent drug use, let alone allow the prediction of possible drug effects. It is therefore advisable to include blood testing if an assessment of possible drug effects is required.

Hair

Hair has long been used to test for exposure to heavy metals, such as arsenic, mercury, and lead. Hair has also proved to be a useful specimen for the analysis of drugs. It is particularly useful for establishing drug use many weeks or months prior to collection.

Drug entry into hair is complicated and is likely to involve a number of processes. Incorporation by entrapment from the blood bathing the growing follicle is a major mechanism, although incorporation through direct contact of mature hair with sweat and/or sebaceous secretions is also a significant source of drug entry.

Because of the ability of hair to directly absorb drug, contamination of hair by direct environmental exposure should also be reasonably excluded if hair results are to be used. For example, nicotine is found in the hair of nonsmokers and cocaine is found in the hair of children of cocaine users. This is arguably the major limitation of this specimen.

The target analytes in hair are predominantly the parent drugs. Cocaine, Δ^9-tetrahydrocannabinol (THC), heroin, and 6-acetylmorphine for heroin, and benzodiazepines are found in higher concentrations than their corresponding metabolites. In this respect, hair concentrations parallel more closely sweat concentrations than those of other specimens.

There are a number of factors that affect retention of drugs into hair: hair colour is one well-known factor. Pigmented hairs, particularly those found in

black-haired Africans and Asians, have higher levels of cocaine than the often weakly pigmented Caucasian counterparts. This is likely to be true for all basic drugs which bind to forms of melanin, the major pigment in hair. Acidic drugs tend to have lower concentrations than basic drugs. Bleaching and the excessive use of shampoo and conditioners can also reduce the concentration of drugs in hair. For this reason, and because of the various routes of drug intake to hair, quantitative results in hair are rarely useful.

Sweat

Sweating is a physiological process providing a mechanism of reducing body temperature. Sweat is produced by eccrine glands, located in the transdermal layer of most skin surfaces and apocrine glands located in axillary and pubic regions. Approximately 50% of all sweat is produced by the trunk, 25% by the legs, and 35% by the head and upper extremities. Sweat is approximately 99% water, the remainder being sodium chloride. A rate of sweating of over 20 ml h^{-1} is common. Sweat glands are often associated with hair follicles, and therefore, it is sometimes difficult to differentiate between the presence of drugs in hair and sweat.

Sweat is normally collected by using suitable absorbent devices such as sweat patches. Contact time may vary from a simple swipe over a portion of skin, to days for a sweat patch to absorb accumulated sweat. The device used and collection time will affect the detection of excreted drugs.

Advantages of sweat analysis are the high tolerability of patches, with correspondingly low incidence of allergic reactions, the ability to monitor drug use over several weeks, and the ability to withstand tampering. Disadvantages include ensuring cleanliness of skin and excluding other environmental contamination.

Drugs detected in sweat include alcohol (ethanol), amphetamines, cocaine, benzodiazepines, barbiturates, opioids, and phencyclidine.

Saliva

Saliva is excreted primarily by three glands – the parotid, submaxillary, and sublingual – and by other small glands such as the labial, buccal, and palatal glands. Mixed saliva used for drug analysis is approximately 65% submandibular, 23% parotid, and 4% sublingual; the remaining 8% is from the other three glands.

The daily flow of saliva in an adult ranges from 500 to 1500 ml. Saliva flow is mediated by a number of physiological factors, paticularly emotional factors, age, gender, and food intake.

Saliva is not an ultrafiltrate of blood, rather it is a complex fluid formed by different mechanisms against a concentration gradient, by pinocytosis, by ultrafiltration through pores in the membrane, and by active transport. Passive diffusion is apparently the dominant mechanism.

Saliva is best collected by absorption onto an absorbent material or a device which stimulates production of saliva. A number of such devices are available to facilitate the collection process. It is also essential that collection of saliva takes place at least 30 min after a meal, or consumption of a beverage or drug, and the oral cavity is free from food material and other objects prior to collection.

The main disadvantage is that saliva volumes are usually small, hence there will be limited ability to repeat analyses. Additionally, not all subjects will be able to provide saliva on demand.

Interpretation of saliva drug concentrations is more difficult than in blood because saliva concentrations are subject to a greater number of variables, such as the degree of protein-binding, pK_a of the drug, and the pH of the saliva. For some drugs (e.g. barbiturates, benzodiazepines), saliva concentrations are much lower than for blood, whereas, for others (e.g. amphetamines), concentrations are higher.

Techniques

A variety of techniques is available for the detection of drugs in specimens collected ante mortem. These range from commercial kit-based immunoassays and traditional thin-layer chromatography (TLC) to sophisticated instrumental separation techniques, such as HPLC, GC, and capillary electrophoresis (CE). MS is the definitive technique to establish proof of structure of an unknown substance, although a number of other detectors can be used to identify the presence of unknown substances in biological specimens.

Immunoassays

A number of different immunoassay methods are available for drugs of abuse. Numerous commercial kits now exist for this purpose. These include enzyme immunoassays (EIA) (e.g. EMIT) and enzyme-linked inmmunosorbent assays (ELISA), fluorescent immunoassays (FPIA) (e.g. Abbott TDx and ADx), agglutination or kinetic interaction of microparticles, immunoassays (e.g. TRIAGE and ONLINE), cloned enzyme donor immunoassay (CEDIA), and radio-

immunoassays (RIA) (DPC assays). These kits also include devices for rapid on-site testing of blood, urine, and sweat without the need for analyzers.

These tests have the advantage of recognizing more than one member of a class of drugs, e.g. amphetamines, benzodiazepines, opioids. However, not all members are detected with equal sensitivity, which will not only be dependent on the crossreactivities of the antibodies to the benzodiazepines, but also on the profile of metabolites present in urine, and the amount of target drug. Different batches of antibody will also influence the sensitivity and selectivity to benzodiazepines and their metabolites.

The overall sensitivity can also be increased by prior hydrolysis of urine to convert glucuronide and sulfate conjugates to substances that are detectable by the kit, although reducing recommended cutoff concentrations can accommodate most of the loss of sensitivity. This technique is particularly useful for cannabis, morphine, and the benzodiazepines that are metabolized to conjugates.

Most kits are directed to urine, although many are available for plasma. Urine-based kits can be used for all types of antemortem specimens by appropriate modification. Precipitation of blood proteins by treatment with methanol, acetonitrile, dimethylformamide, or acetone, and direct analysis of the supernatant are frequently used techniques; however, the high-potency drugs are not always detected. Prior extraction of blood with a solvent (e.g. butyl chloride) provides improved detectability because a concentration step can be employed and most interferences have been removed. With all of these techniques, not all drugs are extracted. Individual validation must be conducted to ensure adequate detectability.

False-positive results with immunoassays occur, either from structurally related drugs, from metabolites of other drugs that are recognized by the antibodies, or occasionally by artifacts, such as adulterants affecting pH, detergents, and other surfactants. For this reason, any positive result must be confirmed by an alternative technique, preferably chromatography with mass spectral identification.

Thin-layer chromatography

This is the oldest of the chromatographic techniques and is still widely used in forensic laboratories as a screening technique. The movement of an organic-based solvent on a plate containing an absorbent material is based on the separation of drugs (and their metabolites). The stationary absorbent phase is typically silica, although other supports are used. Chromatography is usually rapid (less than an hour) and a number of samples can be run simultaneously at little cost. Drugs are identified by visualization under ultraviolet light (as a dark spot), or by spraying with one of a number of reagents, which are directed to specific chemical moieties (as a colored spot), or to organic compounds generally.

The retention factor is calculated by dividing the distance moved from the origin over the distance moved by the solvent front. Characteristic colors of the spots, presence of metabolite patterns and the retention factor values provide a good means of identifying drugs in biological specimens. Unfortunately, the technique is relatively insensitive and is usually limited to urine analysis, although analysis of gastric contents and liver extracts (in postmortem analysis) is also possible. Densitometry of TLC plates can provide some quantitation of the amount of drug present in an extract. Detection limits of $500\,\mathrm{ng\,ml^{-1}}$ are possible from 5 ml of urine.

The use of high-performance TLC plates (HPTLC) has been shown to provide higher sensitivity and can detect some drugs down to $100\,\mathrm{ng\,ml^{-1}}$ from 1 ml of blood. Since specificity is not very high, it is still advisable to confirm any positive result by an alternative technique, preferably MS identification.

High-performance liquid chromatography

HPLC is a commonly used chromatographic system that involves the separation of compounds by partitioning between a pressurized moving liquid phase and a solid support containing very fine silica (4–10 μm diameter particles) or bonded silica. The bonded ligand acts as a pseudoliquid phase. Bonded groups include C2, C8, C18, CN-alkyl, and phenyl-alkyl chains. The physiochemical properties of the bonded phase and the moving phase determine the separation process.

Moving phases are often hydroalcoholic solvent systems, such as methanol/unbuffered water to solvent/buffered phosphate solutions, the base modifier triethylamine and ion-pairing reagents, such as methane sulfonic acid, tetramethyl ammonium hydrogen sulfate, and tetrabutyl ammonium bromide. Gradient programming, in which the composition of solvent is altered with time, provides an ability to separate compounds of widely differing polarity. Normal phase chromatography on a CN-, OH-bonded column or a silica column functions in a similar way to TLC, except that resolution and sensitivity is far higher.

Detection of the sample is most often by ultraviolet spectrophotometry at or near the maximum absorp-

tion wavelength. Alternatively, other physiochemical properties of the compound(s) can be exploited. These include infrared, fluorescence, phosphorescence, electrochemical properties, and conductivity (for ionically charged substances). Compounds with functional groups can be reacted with reagents to impart greater detectability with one or more detectors, or to allow resolution of stereoisomers (**Table 2**).

Photodiode array detection (to supplement ultraviolet light detection) offers real advantages to analysts in identifying peaks and assisting in establishing peak purity. Photodiode array detection can be a very useful technique if MS instrumentation is not readily available, or if absolute proof of structure is not required.

Detection limits around 10–50 $ng\,ml^{-1}$ are expected for most compounds by HPLC, depending on the physiochemical properties of the drug, the volume of specimen extracted, and the method used. Lower detection limits are possible if larger amounts of sample are extracted and when a concentration step is employed.

Solid-phase extraction using small columns to selectively absorb drug from the matrix (e.g. Extrelut, Sep-Pak, Bond-Elut, etc.) provides an excellent alternative to conventional liquid–liquid extraction techniques. Solid-phase techniques have been published for most analytes and tend to be quick, often provide clean extracts, and can be readily automated.

Narrow-bore columns ($\sim$1–2 mm internal diameter) require less specimen and can easily be interfaced with MS. The combination of HPLC with MS (LC-MS) and tandem MS (LC-MS-MS) provide good examples of the separation power of HPLC with the sensitivity and specificity of MS. Detection limits range from 10 pg on-column, resulting in detection limits of better than 1 $ng\,ml^{-1}$ for many compounds using a thermospray or electrospray interface.

Gas chromatography

GC is based on the principle of partitioning a substance in a gaseous phase from a stationary liquid phase. The stationary phase is typically a polymeric liquid, which is either coated on to silica or chemically coated onto the glass surface of the column itself. The nature of the functional groups and polarity of the polymer, and the temperature of the column, provide the means to vary the separation conditions.

Typically, columns are flexible capillaries made of fused silica, with internal diameters of 0.1–0.5 mm, and are coated with heat resistant polymers to promote flexibility. A large range of columns is available to provide analysts with sufficient flexibility to optimize separation conditions. The type of columns range from low polarity dimethylpolysiloxane, 14% cyanopropylphenyl, 5% diphenyl methylpolysiloxane to the polar trifluoropropylpolysiloxane to 50% diphenyl methylpolysiloxane phases. The use of a cyanopropylphenyl or 5% phenylmethylsilicone stationary phases can give better separation of a number of moderately polar compounds than a 100% methylsilicone phase. Due to the wide polarity differences of drugs, temperature programming is necessary for assays involving detection of a number of drugs.

A range of detectors is available for GC. Flame ionization detectors are workhorse detectors for any compounds containing carbon, whereas a number of detectors are available for specific functional groups. The nitrogen–phosphorus detector selectively detects compounds with either nitrogen or phosphorus, while the electron capture detector relies on the ability of a compound to capture electrons when passing through an electric field. Electron capture detectors give the best detection limits ($\leqslant$1 $ng\,ml^{-1}$)

Table 2 Detection systems used in HPLC analysis of selected drugs.

Drug class	*Detector*
Amphetamines	UV and F (of derivatized drug)
Analgesics (acetaminophen, salicylate)	UV and photodiode array
Anions (bromide, chloride, azide, etc.)	Ion conductivity
Antidepressants	UV and photodiode array
Benzodiazepines	UV and photodiode array
β_2-stimulants (salbutamol, fenoterol, etc.)	F
Cannabinoids (THC, carboxy-THC, etc.)	ECD and photodiode array
Catecholamines (epinephrine, dopamine, etc.)	ECD
Cocaine and metabolites	UV and photodiode array
Morphine/codeine	ECD, F and UV
Quinine/quinidine	F and UV

ECD, electrochemical detector; F, fluorescence; THC, Δ^9-tetrahydrocannabinol; UV, ultraviolet.

from 1.0 ml plasma, although the nitrogen–phosphorus detector provides detection limits down to $5\,ng\,ml^{-1}$ for nitrogenous substances, and better than $1\,ng\,ml^{-1}$ for phosphorus-containing substances (e.g. organophosphate pesticides) (**Table 3**). Poisonous and other gases can be detected by use of thermal conductivity detectors which do not rely on the presence of carbon or nitrogen.

For drugs to be amenable to GC they must be thermally stable to enable volatilization into an inert gas (e.g. helium, nitrogen). In many cases compounds can be derivitized to improve their thermal stability, or to alter their retention characteristics and thus, enable a separation to occur (**Table 4**).

Solid-phase microextraction is a relatively recent technique, enabling rapid analysis of drugs without requiring extensive sample clean-up and concentration. Direct online injection using a dialysis technique involving a copolymer precolumn for absorption has also been reported on small sample volumes.

Capillary electrophoresis

A powerful emerging technique showing widespread application in forensic science is that of CE. It is actually a number of related techniques, including capillary zone electrophoresis, micellar electrokinetic capillary chromatography, capillary electrochromatography, capillary isotachophoresis, capillary gel electrophoresis, and capillary isoelectric focusing, and is complementary to HPLC with high separation power.

In its most simple form, CE employs a separation capillary of 20–100 μm internal diameter and up to 100 cm long, a high voltage source, electrodes, an injection system, and a detector. The capillary is often fused silica, coated with plastic polyimide to confer elasticity. The capillary ends are dipped in buffer and are held at a potential of up to 30 kV. The separation is based on migration of charged drug molecules against an electric field, and electroosmosis caused by the osmotic migration of cations and water to the cathode as a result of ionization of the silyl hydroxyl groups on the fused silica. The electroosmosis factor can be altered by changing the pH of the buffer, ionic strength of buffer, modifiers added to buffer, and type of capillary internal wall coating.

The amount of sample or biological extract applied to CE is in the nanogram scale, allowing for trace analysis with adequate sensitivity for most applications.

Electrokinetic micellar chromatography has been shown to be capable of the analysis of illicit drugs in urine and in plasma. This is a powerful technique, as it can separate a large range of compounds with high sensitivity and has the ability to separate compounds of widely differing polarity in one run.

Multiwavelength ultraviolet light detection can be used to provide an added degree of confirmation. The sensitivity is adequate for routine confirmatory analyses of presumptive positive urine specimens for drugs of abuse. CE can also be linked to other detectors, including the mass spectrometer.

Mass spectrometry

MS is the definitive technique if unequivocal identification of unknown compounds is required for forensic purposes. It is usually linked directly to a chromatographic separation process such as CE, HPLC, or GC, or even to another mass spectrometer (MS-MS).

Compounds do not always show characteristic spectral detail (e.g. amphetamines). Consequently, it

Table 3 Detection systems used in GC analysis of selected drugs.

Drug class	*Detector*
Amphetamines	NPD, EI-MS, NCI (as derivative)
Antidepressants	NPD, EI-MS
Antipsychotics	NPD, EI-MS
Benzodiazepines	NPD, ECD, NCI
Cannabinoids (THC, carboxy-THC etc.)	EI-MS, NCI (as derivative)
Carbon monoxide, and other gases	TCD
Cocaine and metabolites	NPD, EI-MS (as derivative of BE)
Heroin, morphine and other opioids	NPD, EI-MS (as derivative for morphine)
Organophosphate pesticides	NPD, EI-MS

BE, benzoylecgonine; ECD, electron capture detector; EI-MS, electron impact mass spectrometry; NCI, negative ion chemical ionization mass spectrometry; NPD, nitrogen–phosphorus detector; TCD, thermal conductivity detector; THC, Δ^9-tetrahydrocannabinol.

Table 4 Derivatives used in GC-MS analysis of selected drugs.

Drug class	*Detector*
Amphetamines	AA, HFBA, methyl chloroformate
Barbiturates	None, or iodomethane in TMAH
Benzodiazepines	*t*-butyl-DMS, TMS, PC/PI
Cannabinoids (THC, carboxy-THC etc.)	TFAA, TMS, PFPA/PFP, *t*-butyl-DMS
Cocaine and metabolites	*t*-butyl-DMS, PFPA/PFP, TMS
Morphine	HFBA, TMS

AA, acetic anyhdride; HFBA, heptafluorobutyric anhydride; PFPA, pentafluoropropionic anhydride; PC, propionyl chloride; PI, propyl iodide; PFP, pentafluoropropan-2-ol; *t*-butyl-DMS, *t*-butyl dimethylsilyl; TFAA, trifluoracetic anhydride; TMAH, trifluoracetic anhydride; TMS, trimethylsilyl;

is recommended that derivatives should be prepared for such compounds, or for substances which show poor chromatographic properties (Table 3). One of the derivatives most frequently described is the trimethylsilyl ether for amines, hydroxy-, and carboxyl-containing substances. Alternatively, other silyl ethers such as *t*-butyl are used, and fluorinated acyl anhydrides (e.g. pentafluoropropionic anhydride) are widely used for amines and hydroxy compounds, and a combination of a perfluorinated alcohol with a perfluorinated acyl anhydride for carboxy-, hydroxy-, and amine-containing substances. Other derivatives are also known.

Positive-ion chemical ionization produces a much higher intensity molecular ion, and is often used to reduce fragmentation and to provide evidence of the molecular weight of the compound. In this mode, reagent gases, such as methane and ammonia, are used to produce different ion–molecule collisions in the ion chamber (source).

The use of negative-ion chemical ionization affords a greatly enhanced detection limit for certain compounds, compared with electron impact mass spectrometry. In this NCI mode a single ion cluster is often observed and can provide, for some drugs (e.g. benzodiazepines and derivitized THC), a detection limit of $0.1\,\mathrm{ng\,ml^{-1}}$.

The use of deuterated internal standards provides an ideal way of monitoring changes in chromatographic performance, and, most importantly, essentially eliminating matrix effects caused by poor recoveries of drug. While recoveries of drug may vary from one matrix to another, and even from calibrators, the deuterated internal standard will correct for this. For this reason, assays involving MS should use deuterated internal standards wherever possible.

See also: **Toxicology**: Methods of Analysis – Post Mortem.

Further Reading

Backer DR (ed.) (1995) *Capillary Electrophoresis*. New York: Wiley.

Davies G (ed.) (1986) *Forensic Science*. Washington, DC: American Chemical Society.

De Zeeuw RA (1989) Modern chromatographic procedures in systematic toxicological analysis. *Journal of Chromatography* 488:199–213.

De Zeeuw RA *et al.* (1995) *Hair Analysis in Forensic Toxicology*. Abu Dhabi: Directorate of Police.

DFG/TIAFT (1992) *Thin-layer Chromatographic Rf-values of Toxicologically Relevant Substances in Standardized Systems*. Weinheim: VCH.

DFG/TIAFT (1992) *Gas-chromatographic Retention Indices of Toxicologically Relevant Substances on Packed or Capillary Columns with Dimethylsilicone Stationary Phases*. Weinheim: VCH.

Drummer OH (1998) Adverse drug reactions. In: Selby H (ed.) *The Inquest Handbook*. Leichhardt, NSW: Federation Press.

Drummer OH (1999) Review: Chromatographic screening techniques in systematic toxicological analysis. *Journal of Chromatography* 733:27–45.

Ghosh MK (1992) *HPLC Methods on Drug Analysis*. New York: Springer.

Höld KM, de Boer D, Zuidema J and Maes RAA (1996) Saliva as an analytical tool in toxicology. *International Journal of Drug Testing* 1:1–29.

Maurer HH (1992) Sytematic toxicological analysis of drugs and their metabolites by gas chromatography-mass spectrometry. *Journal of Chromatography* 118:3–42.

Moffatt AC (1986) *Clarke's Isolation and Identification of Drugs*. London: Pharmaceutical Press.

Snyder LR, Kirkland JJ and Glajch JL (1997) *Practical HPLC Method Development*, 2nd edn. New York: Wiley.

United Nations (1997) *Recommended Methods for the Detection and Assay of Barbiturates and Benzodiazepines in Biological Specimens*, ST/NAR/28. New York: United Nations.

United Nations (1998) *Guidelines for Testing Drugs Under International Control in Hair, Sweat and Saliva*. Vienna: United Nations.

Methods of Analysis – Post Mortem

O H Drummer, Victorian Institute of Forensic Medicine, Monash University, Melbourne, Australia

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0545

Introduction

Drug detection in postmortem specimens has many uses, all of which assist the investigating authorities in providing the relevant information pertaining to a case (**Table 1**). Ultimately, toxicology testing results will assist the coroner, medical examiner or equivalent in establishing evidence of drug use, or by refuting the use of relevant drugs. This latter is important because a pathological examination will often not show indicia suggestive of drug use. Drug use can only be confirmed by appropriate toxicology testing procedures. Clearly, in cases of sudden and unexpected death, use of drugs may provide a cause of death, or at least provide evidence of drug misuse or drug abuse. ('Drug misuse' is used here to define cases where inappropriate doses or drug combinations have been used, whereas 'drug abuse' is restricted to deliberate and usually recreational use of drugs of abuse. Neither term necessarily implies suicidal intent.)

Toxicology testing is particularly important in victims of homicide, to whom drugs may have been given by the assailant to reduce consciousness, and in cases in which drugs were used by the victim. In this latter scenario, modification of behavior by drug use may be important in criminal trials, not necessarily to mitigate the intent by the accused, but primarily to reconstruct, as far as possible, the events that led to the act. Such reconstruction may involve corroboration of witness accounts of drug-using behavior.

Typical drugs used in these cases are alcohol, amphetamines, cocaine or one of the benzodiazepines (alprazolam, diazepam, flunitrazepam, etc.). Victims or perpetrators of violent crime may also have consumed medication to treat a psychiatric problem or a host of other medical conditions. The presence of drugs may therefore be indicative of such treatment, or at least confirm the person concerned has taken the medication. In some cases these medications may even have contributed to behavioral problems.

Table 1 Reasons for drug testing in postmortem cases.

Establishing drug use in victims of homicide
Establishing drug use in drivers of motor vehicles
Establishing drug use in persons involved in workplace accidents
Establishing drug use in other cases of sudden and unexpected death
Assisting investigators with estimation of timing of drug ingestion

In practice, deceased persons have often consumed two or more drugs, and in many cases the investigating authority (pathologist, coroner, etc.) is not aware of all the drugs being used. Since the great majority of cases ($>70\%$) involve more than one drug, it is advisable to conduct a broad drug screen to include most of the common drugs, rather than target the analysis to one or a limited range of drugs suggested by the circumstances. This will also allow experts to determine if any adverse drug interactions have occurred.

Specimens

The preferred specimens collected at postmortem will depend on the type of case. Most typically, one or more blood specimens and urine are collected, although as **Table 2** illustrates, a number of other specimens should be taken in certain case types. A useful forensic technical procedure in the autopsy suite is to take a 'full' set of specimens in all but the most obvious natural cause investigations. This will avoid the embarrassment of collecting insufficient or inappropriate specimens and give the toxicologist the best chance to complete the analytical investigation satisfactorily.

Blood is the most useful specimen that can be collected because drugs present in this fluid can best be related to a physiological effect and can be used to assess the likelihood of recent drug use or exposure to chemicals. A number of problems are associated with the collection of this fluid in cadavers. Two 10 ml samples of blood are recommended, one to be used for blood alcohol analyses and the other for blood toxicology. The splitting of the two blood specimens reduces the possibility of contamination in the laboratory and enables the blood alcohol specimen to be retained separately from the other blood specimen. Forensic technicians or pathologists should be aware that the collection of peripheral blood reduces the possibility of postmortem artifacts frustrating interpretation of any positive results (see Redistribution). The preferred collection site is the femoral region (leg); however, failure to ligate the femoral vessels can allow 'contaminated' abdominal blood to be withdrawn. Autopsy procedures should therefore accomodate these problems. Drug content is reported as milligrams per liter.

Urine is the second most important specimen collected. Since concentrations of drugs and metabolites of drugs are usually much higher than in blood, urine

Table 2 Recommended minimum specimens to be collected post mortem.

Type of case	*Recommended specimens collected*
All cases	Peripheral blood (2 × 10 ml) Urine (10 ml) Vitreous humor (2–5 ml)
Homicides and suspicious cases	Liver, bile
Drug-related cases	Gastric contents, bile, liver
Volatile substance abuse cases	Lung fluid or tied-off lung, liver
Biochemical abnormalities (insulin, etc.)	Serum
Heavy metal poisoning	Liver, hair, kidney

provides a valuable specimen for assessing drug use over the previous day or two. Urine can be collected after the opening of the abdomen, or by direct puncture of the bladder. An autopsy is therefore not necessary for collection of this specimen, or even blood and vitreous humor (see later). Drug content is normally reported as milligrams per liter.

Vitreous humor is an ideal fluid to collect with 'positive' blood because the alcohol content of vitreous humor is very similar to that of blood and can prove useful in excluding putrefactive formation of alcohol in blood and visceral contamination. Vitreous humor is also a useful fluid for a range of drugs, including digoxin and antidepressants, as well as a number of biochemical markers. Since vitreous humor can be easily collected, it is strongly recommended that this specimen is included in a routine, sudden death investigation. Drug content is normally reported as milligrams per liter.

The liver is traditionally a favored tissue for toxicologists because drugs are often found in higher concentrations there than in blood, it can be readily homogenized and, of course, it is the main metabolic organ. A liver sample should be collected in all cases of suspected drug use. A 100 g aliquot is sufficient for most analyses. The right lobe is preferred, as this is least subject to postmortem diffusion of drug from bowel contents and the mesenteric circulation (see Redistribution). Drug content is normally reported as milligrams per kilogram of wet weight tissue.

Gastric content is invaluable in cases of suspected poisoning. The aim of using this specimen is to establish the actual content of drug (or poison) remaining in this organ at death, and it may allow the route of drug administration to be determined. Drug residues can be isolated by direct extraction with methanol, or another solvent, and analyzed by conventional chromatographic techniques. When little or no fluid is present in the stomach, provision of the whole stomach allows the analyst to dissolve any drug adhering to the sides of the walls. Toxicologists should be aware that small quantities of drug will derive from the bile, especially during agonal processes, hence drug content in the stomach does not necessarily imply oral ingestion. Results should be reported as milligrams (total gastric content).

Lung fluid, or tied-off lungs, are recommended in cases of suspected volatile substance abuse. Since quantitative results are rarely interpretable, 'detected' or 'not detected' results alone are usually sufficient (**Table 3**).

Bile is a useful fluid for morphine/heroin detection because biliary concentrations are much higher than in blood. A number of other drugs are also found in bile in relatively high (and therefore more easily detectable) concentrations, including other narcotics, benzodiazepines and glucuronide metabolites. Report bile results as milligrams per liter.

Occasionally other specimens can provide valuable information in a case. Hair can provide a history of drug use, or exposure to chemicals if chronic exposure is thought to occur. Hair can therefore provide evidence of drug use for much longer periods of time than urine. The relation between dose and hair concentration is usually poor, although some comparisons can be made as to the extent of drug use, e.g. regularity of heroin use. Hair concentrations are normally reported as micrograms per gram weight.

Table 3 Useful substance detections in various tissues.

Tissue	*Substances detected*
Blood/urine/liver/hair/gastric contents	All drugs and poisons
Vitreous humor	Alcohol, antidepressants, narcotics, creatinine, urea, glucose ($>10\ \mathrm{mmol\,l^{-1}}$)
Bile	Morphine and other narcotics, benzodiazepines
Lungs	Volatile substances (toluene and other solvents, butane and other aerosol gases, automobile and aviation fuels)

Samples of brain tissue may be more relevant for some centrally active (i.e. in the brain and spinal cord) drugs, such as morphine; and skin (with associated subcutaneous tissue) may show large deposits of drugs left behind after an injection. Results are normally expressed as milligrams per gram wet weight tissue.

In cases of extreme putrefaction, the recommended list of specimens will no longer be appropriate. Muscle tissue, hair and bone can be useful specimens in this type of case, although the physical state of the body will determine what specimens are available for collection. Body fluids will be present in some putrefied bodies but this is no longer blood, rather liquified tissues; but this fluid can be used to screen for the presence of drugs. Quantitative results are of little use in badly putrefied cases.

General Techniques

The range of techniques available for the detection of drugs in the specimens collected post mortem are essentially identical to those collected ante mortem. These range from commercial kit-based immunoassays (enzyme multiplied immunoassay technique (EMIT), fluorescence polarization immunoassay (FPIA), cloned enzyme donor immunoassay (CEDIA), radioimmunoassay (RIA), etc.), traditional thin-layer chromatography (TLC), to instrumental separation techniques such as high-performance liquid chromatography (HPLC), gas chromatography (GC) and capillary electrophoresis (CE). Mass spectrometry (MS) is the definitive technique for establishing proof of structure of an unknown substance and can be linked to GC, HPLC and more recently to CE.

The specimens analyzed in postmortem cases are most often blood and liver, rather than urine and serum, which are used in antemortem analysis, and the other specimens listed earlier. The use of blood and liver, and indeed all other postmortem specimens, require separate validation to those methods used in antemortem analysis. The methods used require modification to ensure a reliable extraction recovery, a low level of interference and reproducible quantitative results. Special attention to these factors is required on partly or fully putrefied specimens to ensure lack of interference from endogenous substances. Cutoff values often used in workplace, sports and drugs-of-abuse testing are no longer appropriate in postmortem cases involving specimens other than urine. Even postmortem urine should not normally be tested to cutoff limits used in drugs-of-abuse testing because the presence of a small concentration of drug may be of forensic significance.

The range of immunoassays used in antemortem analysis can also be used in postmortem analysis, providing suitable modification of the preparation of the specimen is conducted. Urine-based kits can be used for urinalysis; however, blood or tissue homogenates require special treatment to remove matrix effects. Urine is often unavailable in posmortem cases. Precipitation of blood proteins by treatment with methanol, acetonitrile, dimethylformamide or acetone, and direct analysis of the supernatant are frequently used techniques; however, the high-potency drugs are not always detected. Prior extraction of blood or liver homogenates with a solvent (e.g. butyl chloride) provides an improved detectability, as a concentration step can be employed and most interference has been removed. Not all drugs are extracted with all of these techniques. Individual validation must be conducted to ensure adequate detectability for target drugs and to determine the range of drugs that can, and those that cannot, be detected.

False-positive results with immunoassays occur, either from structurally related drugs, or from metabolites of other drugs that are recognized by the antibodies. While HPLC and GC techniques are more specific than immunoassays, any positive result should be confirmed by mass spectral identification, unless sufficient validation of another method has been conducted to assure courts of the reliability of the result. Unconfirmed drug results, if reported, should be flagged as presumptive, or words with similar intent.

Solid-phase extraction, using small columns to absorb drug selectively from the matrix (e.g. Extrelut, Sep-Pak, Bond-Elut, etc.), provides an excellent alternative to conventional liquid–liquid extraction techniques. Solid-phase techniques have been published for most analytes, tend to be quick, often provide clean extracts and can be readily automated.

The use of deuterated internal standards provides an ideal way to monitor changes in chromatographic performance, and, most importantly, essentially eliminates matrix effects caused by poor recoveries of drug. While recoveries of drug may vary from one matrix to another, and even between calibrators, the deuterated internal standard will correct for this. For this reason, assays involving MS should use deuterated internal standards wherever possible in postmortem analyses.

Recommended Techniques for Postmortem Analysis

As indicated before, it is important that a drug screen encompass the widest number of drugs and poisons without seriously compromising the ability of the

laboratory to work on sufficient cases. Urinanalysis (or blood or another fluid) using one of the commercial immunoassays, or even TLC, is recommended for the main classes of drugs. These usually include amphetamines, barbiturates, benzodiazepines, cannabinoids, cocaine metabolites and morphine-like opiates.

In addition, a series of other (usually chromatographic) tests is strongly recommended. The schema shown in **Fig. 1** illustrates a typical analytical profile for routine case screening of blood. Blood is analyzed for alcohol and is subject to a few screening techniques aimed at capturing a wide selection of chemical substances to which humans are likely to be exposed.

An acidic screen includes the nonnarcotic analgesics (acetaminophen, aspirin), the nonsteroidal antiinflammatory drugs (naproxen, ketoprofen, ibuprofen, etc.), many of the diuretics (frusemide, hydrochlorothiazide, etc.), the anticonvulsants (carbamazepine, lamotrigine, phenobarbital, phenytoin, valproate), barbiturates and the more potent benzodiazepines, the xanthines, theophylline and caffeine. The use of a solvent extraction technique at acidic pH, or simple precipitation of blood proteins with acetonitrile, enables these substances to be detected by gradient HPLC with multiwavelength or photodiode array detection.

A basic extraction procedure using butyl chloride (preferred solvent, but others are also suitable), or a solid-phase extraction procedure using octadecyl-bonded cartridges or mixed-phase cartridges, will provide a reasonably clean extract from postmortem blood (and other tissues) for analysis by capillary GC. The use of an MS detector is preferred (to allow simultaneous detection and confirmation), although a nitrogen–phosphorus detector (NPD) will provide a

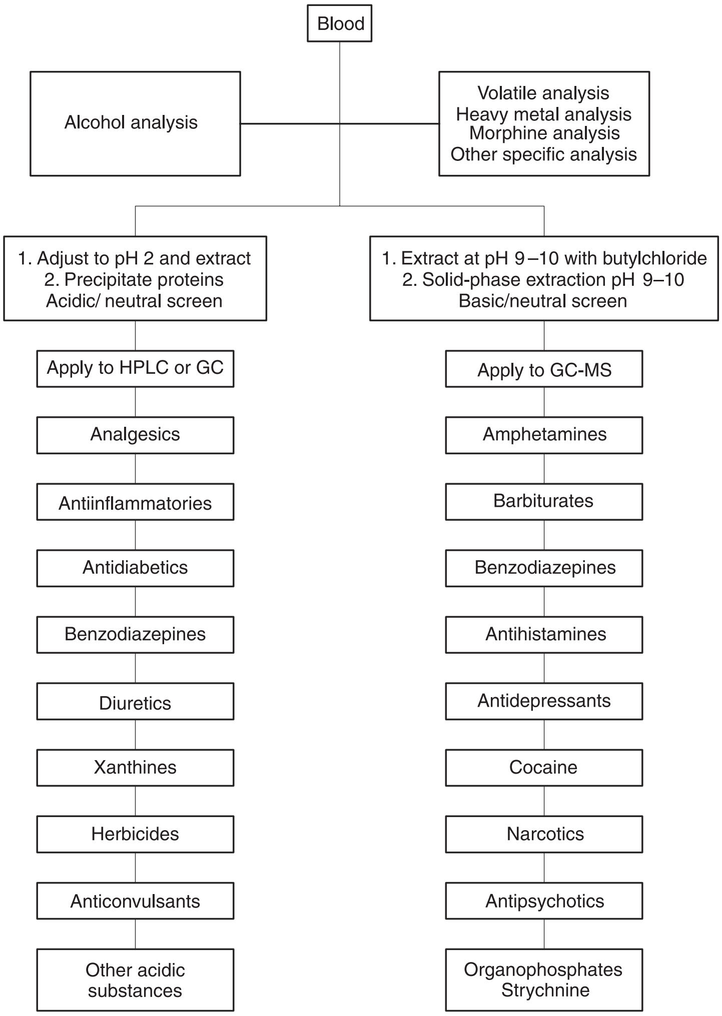


Figure 1 Extraction steps for blood analyses and substance classes likely to be detected. GC, gas chromatography; GC-MS, GC mass spectrometry; HPLC, high-performance liquid chromatography.

higher sensitivity for many substances than full-scan MS. Electron capture detectors (ECD) are extremely useful for benzodiazepines. The use of dual detectors (NPD and MS, or NPD and ECD) provide an additional degree of specificity and detection over one detector alone.

These two screening procedures will also enable a number of unusual poisons to be detected. Organophosphates and strychnine are readily detected by GC-NPD, while HPLC of acid extracts enables detection of a number of herbicides and other agricultural chemicals. If circumstances suggest volatile substance abuse, exposure to heavy metals, lysergic acid diethylamide (LSD) and other nonamphetamine hallucinogens, or other noxious substances not covered earlier, specific additional tests will need to be performed. It is advisable to perform a blood test for morphine if heroin or morphine use is suspected (or needs to be ruled out) and the urine test for opiates is negative, as morphine does not chromatograph well underivatized. Heroin deaths have been missed if screening for morphine is restricted to urine because acute deaths in naive users may not show morphine in urine.

Postmortem Artifacts in Analysis

Dying imparts a number of special processes that affect the collection and analysis of specimens obtained at autopsy.

Redistribution

Foremost is the process of redistribution, which affects all analyses in which concentrations of drugs in blood and tissues alter due to disruption of cellular membranes, causing alterations of drug concentrations within tissue elements and diffusion from one tissue to another. This process is particularly significant for drugs with high lipid solubility, as these drugs tend to show concentration differences in tissues and blood. **Table 4** shows the extent of these changes for selected drugs when comparisons are made between blood collected from the heart and that collected from the femoral region.

The femoral blood is least subject to redistribution after death; however, drugs with much higher concentrations in muscle tissue will still diffuse through the vessel walls and elevate the neighboring blood concentrations. If the femoral vessels are not tied off from the vena cava and aorta, then the process of drawing blood can also extract blood from the abdominal cavity that has been contaminated from diffusion of gastric and intestinal contents. It is therefore advisable to reduce these processes by collecting blood specimens as soon as possible after death from the femoral region, with blood vessels tied off to reduce contamination.

Table 4 Likely extent of postmortem redistribution for selected drugs.

Drug/drug class	*Redistribution*[a]
Acetaminophen	Low
Alcohol (ethanol)	Low
Barbiturates	Low to moderate
Benzodiazepines	Low to moderate
Cocaine	Low
Digoxin	Very high
Methadone	Moderate
Morphine	Low
Phenothiazines	Moderate to high
Propoxyphene	Very high
Salicylate	Low
Serotonin reuptake inhibitors	Low to moderate
Tricyclic antidepressants	High

[a] Low, up to 20% elevation; moderate, 21–50%; high, 50–200%; very high, >200%.

These processes are not limited to blood. Liver and lung tissue show differences in the concentration of drugs, depending on the nature of the drug and whether diffusion of drug has occurred from neighboring tissues or the blood supply. For example, the left lobe of the liver is more likely to exhibit elevated drug concentrations than the right lobe.

Metabolism and bioconversion

A number of drugs can undergo chemical changes in the body after death. These chemical changes can be either metabolically mediated or caused by spontaneous degradative processes. For example, the metabolism of heroin to morphine occurs in life and also occurs in recently deceased persons by the action of blood and liver esterases. For this reason, heroin or the intermediate 6-monoacetyl morphine are rarely detected in cadaveric tissues. Morphine is therefore the target drug. Aspirin is converted rapidly to salicylate by hydrolytic mechanisms. Most prodrugs activated by de-esterification or hydrolysis will be subject to similar processes.

Nitro-containing drugs, such as the benzodiazepines, nitrazepam, clonazepam, nimetazepam, flunitrazepam and others, are also rapidly biotransformed after death to their respective amino metabolites by the action of certain types of bacteria (obligate anaerobes). Toxicologists must therefore target their analyses to these transformation products rather than the parent drug.

Sulfur-containing drugs, such as dothiepin, thiopental, thioridazine, etc., are also subject to bacterial attack post mortem, leading to progressive losses over time during putrefaction. Of course, the parallel process of tissue loss will also affect the tissue concentration during putrefaction.

Chemical instability occurs in a number of drugs and metabolites even when specimens are stored frozen at $-20^{\circ}C$. Some benzodiazepines and benzodiazepine metabolites, antipsychotics such as thioridazine, and the β-stimulant fenoterol show time-dependent losses. Stability characteristics have not yet been evaluated for many drugs.

Alcohol will be lost to evaporation unless sealed tubes are stored at $-60^{\circ}C$, however, alcohol can also be produced by bacterial action on glucose and other sugars found in blood. The use of potassium fluoride as preservative (minimum 1% w/v) is required to prevent bacterial activity for up to 1 month after collection, when stored at 4°C.

Reports

Once an analysis is complete, a report that accurately details the analytical findings must be issued to the client(s). These results should indicate the type of tests conducted, the analytical method used (i.e. HPLC, GC–MS, etc.), on which specimens the analyses were conducted, and, of course, the result(s). The result(s) should be unambiguous, using such terms as 'detected' or 'not detected'. The use of the term 'not present' should be avoided, as it implys no possiblity of the substance being present. A toxicologist can rarely be so definite and can only indicate that a substance was not detected at a certain threshold concentration. A detection limit should therefore be provided when tests for specific substances produce 'not detected' results.

For quantitative results, consistency in units is advised and should not be given with more significant digits than the accuracy will allow. For example, there is no point in reporting a result for blood morphine as $0.162\,mg\,l^{-1}$ when the accuracy and precision of the method is $\pm 20\%$. A result of $0.16\,mg\,l^{-1}$ would suffice.

For drug screening results, it is advisable to provide clients with an indication of the range of substances a method is capable of detecting, and some indication of the detection limits, i.e. 'at least therapeutic concentrations' or 'only supratherapeutic concentrations'. Positive immunoassay results should also be reported even if this presumptive detection has not been confirmed. This information can be useful because it may imply (to an expert later investigating the case) that the substance may have been present but at very low concentrations, or there was another immunoreactive compound which was not excluded in the confirmation assay. To exlude these results could be construed by courts as a deliberate withholding of evidence.

To enable proper interpretation of evidence, all reports should indicate the site of blood sampling, and provide, where relevant, some comment on the possibility of postmortem artifacts, such as redistribution. By incorporating these comments, uninformed persons reading the report are less likely to unwittingly misinterpret the results.

See also: **Toxicology**: Methods of Analysis – Ante Mortem.

Further Reading

Backer DR (ed.) (1995) *Capillary Electrophoresis*. New York: Wiley.

Davies G (ed.) (1986) *Forensic Science*, 2nd edn Washington, DC: American Chemical Society.

De Zeeuw RA (1989) Modern chromatographic procedures in systematic toxicological analysis. *Journal of Chromatography* **488**:199–213.

De Zeeuw RA (1995) *Hair Analysis in Forensic Toxicology*. Abu Dhabi: Directorate of Police.

DFG/TIAFT (1992) *Thin-layer Chromatographic Rf-values of Toxicologically Relevant Substances in Standardized Systems*. Weinheim: VCH.

DFG/TIAFT (1992) *Gas-chromatographic Retention Indices of Toxicologically Relevant Substances on Packed or Capillary Columns with Dimethylsilicone Stationary Phases*. Weinheim: VCH.

Drummer OH (1998) Adverse drug reactions. In: Selby H (ed.) *The Inquest Handbook*, Leichhardt, NSW: Federation Press.

Drummer OH (1999) Review: Chromotographic screening techniques in systematic toxicological analysis. *Journal of Chromatography* **733**:27–45.

Ghosh MK (1992) *HPLC Methods on Drug Analysis*. New York: Springer.

Li SFY (ed.) (1986) *Capillary Electrophoresis, Principles, Practice and Applications*. Amsterdam: Elsevier.

Maurer HH (1992) Systematic toxicological analysis of drugs and their metabolites by gas chromatography–mass spectrometry. *Journal of Chromatography* **118**:3–42.

Moffatt AC (ed.) (1986) *Clarke's Isolation and Identification of Drugs*. London: Pharmaceutical Press.

Pfleger K, Maurer HH and Weber G (1992) *Mass Spectral and GC Data of Drugs, Poisons and their Metabolites*. Weinheim: VCH.

United Nations (1997) *Recommended Methods for the Detection and Assay of Barbiturates and Benzodiazepines in Biological Specimens*. Publication ST/NAR/28. New York: United Nations.

V

Visible Spectroscopy *see* **Analytical Techniques**: Spectroscopic Techniques; Spectroscopy: Basic Principles.

VOICE ANALYSIS

D Meuwly, Institut de Police Scientifique et de Criminologie, Lausanne, Switzerland

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0617

Introduction

In Stanley Kubrick's cult movie *2001, A Space Odyssey* (1968), the hero goes through customs on presentation of his voice. Science fiction and daily experience of safely recognizing people by their voices let us suppose that speaker recognition is a resolved problem, both for human being and for machines.

The human being has a good ability to identify familiar voices and the results of automatic speaker recognition are fair in controlled conditions, but these circumstances are rarely reached in real forensic cases. The speaker recorded from the telephone line in a kidnapping or blackmail case is often stressed or afraid, and unfamiliar to the police. Drug dealers are regularly wiretapped, but because of their nomadic, illegal activity, and as they suspect they are listened to by the police, they use phone boxes or cellular telephones in noisy environments, like cars, streets or bars.

Everyone knows of forensic speaker recognition, but the main areas of forensic science that include voice are: (1) speaker recognition, (2) speaker profiling, (3) intelligibility enhancement of audio recordings, (4) transcription and analysis of disputed utterances, and (5) authenticity or integrity examination of audio recordings. As a primary investigation, authenticity or integrity examination of audio recordings can precede each of the other tasks (1) to (4).

The Human Voice

Nature of voice and production of speech

Voice results from an expiratory energy used to generate noises and/or to move the vocal cords, which generate voiced sounds. This behavior is one of the basic methods of communicating by common codes; these codes are languages.

Speech production is composed of two basic mechanical functions: phonation and articulation. Phonation is the production of an acoustic signal. Articulation includes the modulation of the acoustic signal by the articulators, mainly the lips, the tongue and the soft palate, and its resonance in the supraglottic cavities, oral and/or nasal. The phonemes produced are divided into voiced and unvoiced consonants and vowels; they can be characterized in the time domain, in the spectral domain and in the time-spectral domain.

The frequency range of the normal speech signal is 80–8000 Hz, with a dynamic range of 60–70 dB. The average fundamental frequency of vibration of the vocal cords (F_0), called pitch, is between 180 and 300 Hz for females, between 300 and 600 Hz for children and between 90 and 140 Hz for males.

Perception of voice and speech

Speech perception is generally described as a five-stage transformation of the speech signal in a message: peripheral auditory analysis, central auditory analysis, acoustic-phonetic analysis, phonological analysis and higher order analysis (lexical, syntactic and semantic). The human ear is primarily designed to perceive the human voice. The accepted range for perception is between 16 and 20 000 Hz, with

extremely good sensitivity between 500 and 4000 Hz. The recognized limit in the intensity domain is between 130 and 140 dB.

Voice as Evidence

Collection of evidence

The evidence does not consist of speech itself, but of a transposition, obtained by a transducer, that converts acoustic energy into another form of energy: mechanical, electrical or magnetic. This transposition is recorded on a storage medium, on which it is coded by an analog or digital information-coding method.

In analog recording, the strength and shape of the transduced audio waveform bears a direct relationship to the original sound wave. By contrast, in a digital recording, the transduced waveform is sampled, and each sample is translated into a binary number code, such as that used by computers. **Table 1** lists a variety of audio recording devices (but certainly not all), according to the type of storage medium and coding method used.

Quality of Evidence

Types of evidence

In most cases, evidence is recorded off the telephone lines, either in cases of wiretapping or anonymous calls. In anonymous calls, the message is generally short, from seconds to minutes; where it is a monolog, it may be a prerecorded message, voluntarily modified by a filtering or editing procedure. From a lexical point of view, the themes are targeted, i.e. abuse, extortion, obscenity, threats. In wiretapping procedures, the recordings can reach hundreds of hours. There are conversations, diversified from a lexical point of view, and some utterances may refer to internal codes of groups or organizations.

Speaker variability and simulation

Disputed utterance When making an anonymous call, the voice can be unintentionally altered by the speaker because of the peculiar psychological conditions of stress and/or fear created by the act of committing a crime. State of health, tobacco and psychotropic substances also affect the voice. The speaker may also deliberately modify the voice, speech and/or language.

In wiretapping procedures, speakers are generally unaware of the fact that they are being recorded. In such cases, there are usually no unintentional or deliberate alterations of voice, speech and/or language, as described above. But the spontaneity pursuant to this ignorance leaves the speakers with a substantial adaptability of their discourse in the context, in their mood and in the relationship they establish with their interlocutor.

Control recording The control should be recorded in similar conditions to the disputed utterance and should be as representative as possible of the speaker.

Table 1 Audio recording devices

	Magnetic tape	*Tapeless magnetic*	*Non-magnetic*
Analog	Micro cassette recorder Compact cassette recorder Reel-to-reel recorder Eight-track player Camcorder: analog audio component of video tape Answering machines	Wire recorder (obsolete)	Phonograph using: – wax or tin foil cylinder or disc (obsolete) – shellack disk (obsolete) – vinyl record (increasingly obsolete) – motion picture camera: film audio track (nearly obsolete with respect to home use)
Digital	Digital audio tape recorder (DAT) Digital compact cassette recorder (DCC) Stationary head recorders (e.g., DASH recorders) Camcorder: digital audio component of videotape Computer using tape drive backup: helical scan unit (e.g. DAT drive) or stationary head unit	Sampler, computer or digital audio workstation using: – floppy disk – hard disk – removable ‘hard’ cartridge – magneto-optical cartridge	Audio recorder, sampler, computer or answering machine using semiconductor memory (memory chips) Optical laser unit using: – CD (compact disc), CD-ROM (read only memory), CD-I (interactive) – CD-R (recordable) – CD-RW (rewritable) – DVD (digital versatile disk) – Mini disk

Nevertheless, this procedure is often made months or years after the disputed utterance and the human voice changes with time. Moreover, the suspect can be uncooperative and it could prove difficult to get the suspect to simulate the supposed psychological conditions of the disputed utterance. Such circumstances can prevent or restrict effective speaker recognition procedures.

Transmission channel distortion

On telephone lines, the speech bandwidth is reduced to the frequency range between 300 and 3400 Hz: some high frequency speaker-dependent features, as well as pitch, are not transmitted. The auditor perceives the pitch as being all the same, through the integer multiples of F_0, called harmonics.

The dynamic range of the telephone line is only from 30 to 40 dB, the signal suffers from nonlinear distortion and additive noise is picked up and transmitted by the handset microphone. This implies a number of other potential degradations of the speech signal, depending on the characteristics of the microphone (**Fig. 1**) and the telephone network (**Fig. 2**). This network can be fixed or mobile, and the transmission and coding system of the information can be analog, digital (mainly in developed countries) or mixed. The quality of digital networks depends particularly on the coding algorithms and the transmission rate for information.

Recording system distortion

The recording equipment is the only link in the chain that it is possible to control, but it is unfortunately often the weakest. Either the device is not suited to the task or it is improperly maintained or used, which can prevent any useful analysis of the voice. Even if the transmission is digitized, the recordings for law enforcement are still exclusively realized on analog recorders at low tape speed and with correspondingly low quality.

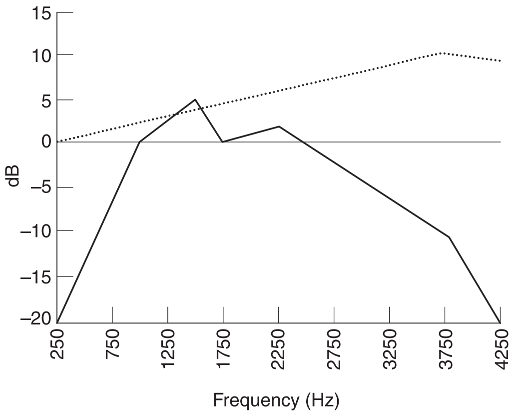


Figure 1 Comparison of frequency responses for two types of handset microphone: electret (····); carbon (—).

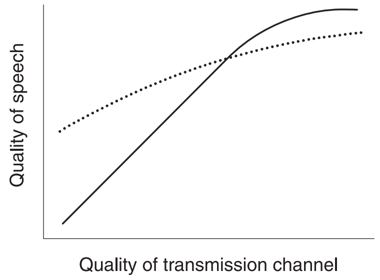


Figure 2 Comparison of transmission quality for two types of mobile network: digital (····); analog (—).

Admissibility

As the conditions of admissibility for audio recording evidence depend dramatically on the criminal law of each country, this topic is not treated here.

Speaker Recognition

Types of speaker recognition

Speaker recognition refers to any process that uses some features of the speech signal to determine if a particular person is the speaker of a given utterance. Three kinds of approach can be distinguished: (1) *s*peaker *r*ecognition by *l*istening (SRL), (2) *s*peaker *r*ecognition by visual comparison of *s*pectrograms (SRS) and (3) *a*utomatic *s*peaker *r*ecognition (ASR).

SRL involves the study of how human listeners associate a particular voice with a particular individual or group, and indeed to what extent such a task can be performed. SRS comprises efforts to make decisions on the identity or nonidentity of voice based on visual examination of speech spectrograms. ASR relies on computer methods, based on information theory, pattern recognition and artificial intelligence.

Each of the above approaches is currently used in forensic science: SRL, practiced either by experts, phoneticians or speech scientists, or by nonexperts, principally victims or witnesses; SRS, practiced by experts; and ASR methods, integrated into semiautomatic or computer-assisted systems. The current tendency is to integrate the results of SRL and ASR and to use spectrography only for visualization purposes (**Fig. 3**); only a few laboratories currently master all approaches.

Procedure and methods

Listening is the initial task. It can also be the only task if the disputed utterance and the control recording obviously differ; for example, if the pitch, the dialect or other grossly audible characteristics of the speaker

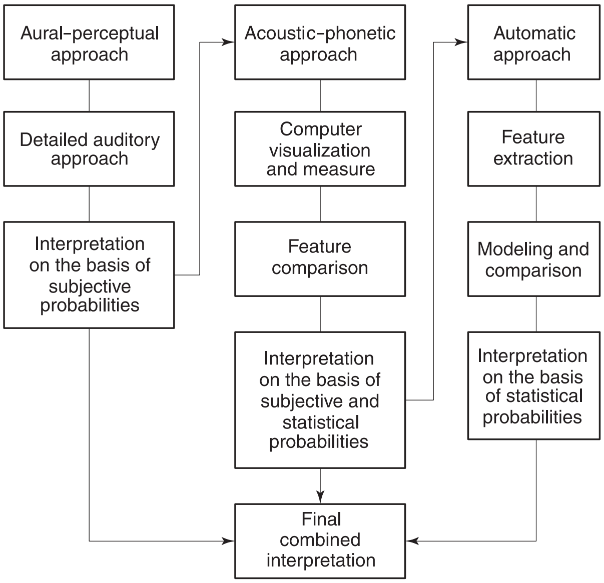


Figure 3 Combined approach for forensic speaker recognition.

are not the same. Therefore, a speaker recognition procedure is more likely to be engaged when close auditory similarities exist between the evidence and the control recording, or when voice disguise is suspected.

When engaged, the recognition process consists of three stages: (1) feature extraction, (2) feature comparison, and (3) classification. Foreign language samples and degraded samples whose signal-to-noise ratio (SNR) is less than 25 dB should be examined only with great caution.

Feature extraction

As no speaker-specific feature is presently known, feature extraction presupposes knowledge of the aspects of the acoustic signal yielding parameters that depend most obviously on the identity of the speaker. Moreover, repetitions of the same material by the same speaker vary from utterance to utterance. The existence of intraspeaker variability makes speaker recognition closely analogous to the identification of handwriting.

The ideal parameters would:

- exhibit a high degree of variation from one speaker to another (high interspeaker variability);
- show consistency throughout the utterances of a single speaker (selectivity);
- preferably be insensitive to emotional state or health and to communication context (low intraspeaker variability);
- withstand attempted disguise or mimicry (resistance);
- occur often in speech (availability);
- be neither lost nor reduced in telephone transmission channel recording process (robustness);
- not be prohibitively difficult to extract (measurability).

Feature comparison

Through her or his utterances, the speaker may be represented in a feature space where the number of dimensions relates to the number of parameters extracted. A distance, appropriate to the parameters, is measured to estimate the closeness of evidence and control. This condition is implicitly accomplished when using a subjective approach. The comparison of the disputed utterance and a control recording leads the scientist either to a numerical assessment, which describes the distance between them, or to a subjective opinion taking into account similarities and differences.

Classification

The numerical value given by ASR expresses a random match probability of the set of features compared, in so far as the range of samples is representative of the population modeled. A subjective opinion, such as given by SRS and an aural–perceptual approach, expresses a subjective probability of the frequency of the set of values. The results of these two last kinds of approach are controversial because a subjective opinion can differ significantly from statistical probabilities, as well as among experts. The acoustic–phonetic approach combines both aspects: numerical and subjective.

Speaker Recognition by Listening (SRL)

Recognition of familiar and unfamiliar speakers are probably two independent and unrelated abilities. Identifying a familiar speaker is essentially a pattern recognition process where a unique voice pattern matches an individual. For discrimination between unfamiliar speakers, the process of matching auditory parameters involves greater feature analysis as well as overall pattern recognition. Human beings perform better identification when the voice is familiar; however, in most cases the perpetrator is unfamiliar to the listener, whether or not expert.

Recognition by nonexperts

Experimental investigations conducted in presenting voice lineups to nonexpert listeners reveal that the accuracy of discrimination of unfamiliar voices depends on several factors: the size and the auditory homogeneity of the voices presented in the voice

lineup; the age and sex of the speakers and the listeners; the quantity of the speech sample initially heard; the delay between the initial hearing of the disputed utterance and the recognition procedure; the existence of a voice disguise; the quality of both the transmission channel and the recording equipment; and the presence or absence of concordant visual testimony.

Even if all these factors are favorable, the great variability of individual performance restricts the use of this type of investigation to indicative value, the uncertainty of which is similar to that of any testimony.

Recognition by experts

This analysis combines aural–perceptual and phonetic–acoustic approaches. Most of the features considered are based upon the physiological mechanisms of the speech production or psychoacoustic knowledge.

Aural–perceptual approach The aural–perceptual approach consists of a detailed auditory analysis. It applies to the parameters (1) of voice, such as height, timbre, fullness; (2) of speech, such as articulation, diction, speech rate, pauses, intonation, speech defects; and (3) of language, such as dynamics, prosody and style. It also includes linguistic observations on syntactic, idiomatic or even paralinguistic features, such as breathing patterns. The main results of this type of analysis are documented in a transcript using the international phonetic alphabet (IPA).

Phonetic–acoustic approach Thanks to computer visualization techniques and specific signal-processing algorithms, the acoustic phonetic examination permits the quantification or the more precise description of several of the parameters studied in auditory analysis. In the spectral domain, the trajectory and the relative amplitude and bandwidth of the formants are studied (the formants, or formant frequencies, are zones of the vowels where the harmonic intensities are focused). The spectral distribution of the energy, pitch and pitch contour are analyzed as well. In the time domain, the study focuses on the duration of the segments, rhythm and cycle-to-cycle variation of the vibration period of the vocal cords, called jitter.

Speaker Recognition by Visual Comparison of Spectrograms (SRS)

Technology

The sound spectrograph is an instrument that shows the variation of the short-term spectrum of the speech wave. In each spectrogram, the horizontal dimension is time, the vertical dimension represents frequency, and the darkness represents intensity on a compressed scale. This speaker recognition technique, widely used in the USA, is based on the assumption that intraspeaker variability is less than or differs from interspeaker variability, and that 'voiceprint' examiners can reliably detect this difference by visual comparison of spectrograms.

The Kersta method

In 1962, Kersta first proposed the SRS method under the name of 'voiceprint' identification. He claimed that the speech spectrogram of a given individual is as permanent and unique as fingerprints and would allow the same level of certainty for forensic identification.

In 1970, Bolt and collaborators denied these assertions, observing that, contrary to fingerprints, the pattern similarity of speech spectrograms depends primarily on acquired movement patterns used to produce language codes, and only partially and indirectly on the anatomical structure of the vocal tract. The details of the pattern are just as variable as the overall pattern and are affected by growth and habits. Therefore, the differences between fingerprints and speech spectrograms are more significant than similarities, and the inherent complexity of spoken language brings SRS closer to aural discrimination of unfamiliar voices than to fingerprint discrimination. The authors concluded that the term 'voiceprint' was a fallacious analogy to fingerprints and that considerable research was necessary to establish the validity and the reliability of the SRS.

The Tosi study: an attempted validation of the Kersta method

In 1972, Tosi and collaborators produced the only large-scale study, to date, for the determination of accuracy for subjects performing speaker identification tasks based on sound spectrograms. They conclude that 'if trained "voiceprint" examiners used listening as well as spectrograms for speaker identification, even under true forensic conditions, they would achieve lower error rates than the experimental subjects had realized under laboratory conditions'.

This extrapolation, as well as the claim that the scientific community had accepted the method, was invalidated by Bolt and collaborators in 1973. They stated that the study of Tosi had improved the understanding of some of the problems of SRS by indicating the influence of several important variables for the accuracy of identification. However, they further concluded that in many practical situations the method lacked an adequate scientific basis for

estimating reliability, and that laboratory evaluations showed an increase in error as the conditions for evaluation moved towards real forensic situations.

The National Academy of Sciences (NAS) report

As a result of the wide controversy on the reliability of the SRS and the admissibility of the resulting testimony in the USA, the NAS authorized the conduct of a study, in 1976, at the request of the Federal Bureau of Investigation (FBI). The committee concluded that the technical uncertainties concerning the SRS were significant and that forensic applications should be allowed with the utmost caution. It took no position for or against forensic use, but recommended that, when in testimony, the limitations of the method should be clearly and thoroughly explained to the factfinder, whether judge or jury.

Since the publication of the NAS report, mixed legal activity has been seen in the USA, some cases favoring admissibility and others disallowing SRS. Since the Daubert decision calling for a 'reliability standard', opponents of SRS might call for disallowance in jurisdictions where it had previously been admitted. Significantly, Daubert is binding only on federal jurisdictions, and the long-term impact of the cases on the USA evidentiary regimes is unknown.

Automatic Speaker Recognition (ASR)

Automatic speaker recognition methods can be divided into text-dependent and text-independent methods. Text-independent methods are predominant in forensic applications where predetermined key words cannot be used.

Feature extraction

During the 1970s, research was devoted to discovering, in the speech signal, the determining factors taking place in human speaker recognition, and to selecting statistically the most relevant parameters. Several semiautomatic systems based on the extraction of explicit acoustic phonetic events and basic statistical measures were developed for law enforcement. SASIS (*s*emi*a*utomatic *s*peaker *i*dentification *s*ystem), developed by Rockwell International in the USA, and AUROS (*a*utomatic *r*ecognition *o*f *s*peaker by computer), developed by Philips GmbH and the BundesKriminalAmt (BKA) in Germany, have been abandoned because of their lack of results in the forensic context or the difficulty in applying them (only usable by phoneticians). SAUSI (*s*emi*a*utomatic *s*peaker *i*dentification system) is under development in the University of Florida, but its applicability to ordinary forensic cases remains to be established.

As early as 1980, the methods based on explicit localization of acoustic events in the speech signal came into use and a consensus has since emerged on the use of derived parameters of the short-term spectrum of the speech signal, such as linear prediction coefficients (LPC) and cepstral coefficients (CC). CAVIS (*c*omputer *a*ssisted *v*oice *i*dentification *s*ystem) was developed at the Los Angeles County Sheriff's Department from 1985. It is based on the extraction of temporal and spectral features and on a statistical weighting and comparison procedure. CAVIS was discontinued in 1989 because the laboratory results, although encouraging, did not match those performed under forensic conditions.

Feature comparison

At present, research focuses on methods able to exploit efficiently whole sets of parameters for the comparison of the control recording and the disputed utterance. Thus, the advances are mainly due to improvement in techniques for making speaker-dependent feature measures and models and do not derive from a better understanding of speaker characteristics or their means of extraction.

In forensic applications, the control recording is used as training data to model the speaker, and the disputed utterance is used as the test for the comparison. Different approaches have been used to model the LPC and CC parameters, principally vector quantization (VQ), ergodic hidden Markov models (HMMs), artificial neural networks (ANNs) and autoregressive vector models (ARVs). The most recent evaluation of the methods realized by the National Institute of Standards (NIST) in the USA shows that the performance of gaussian mixture models (GMMs) surpasses the results of all other methods for text-independent applications.

Normalization techniques

The most significant factor affecting ASR performance is the variation in the signal characteristics from sample to sample, those due to the speaker, those due to transmission and recording conditions, and those due to background noise. Training conditions, test segment duration, sex and handset variation between training and test data affect performance particularly, just as all specific forensic degradations, as cited above. Spectral equalization has been confirmed to be effective in reducing linear channel effects and long-term spectral variation, but normalization can be improved using more recent techniques, such as spectral subtraction combined with statistical missing-feature compensation. However, the current lack of consistency in the algorithms

and the low quality of the audio recording evidence restrict the use of ASR for forensic purposes.

Interpretation of the Results

Forensic identification

The identification process seeks individualization in forensic science. This individualization process can be seen as a reduction process, beginning with the entire suspect population and ending with a single person. In forensic speaker recognition the suspect population can be defined as the population speaking the language of the evidence. The reduction factor depends upon the closeness of the disputed utterance to the control recording, measured by several methods as discussed above. To identify a speaker means that the chance of observing, in the suspect population, an utterance from another speaker presenting the same closeness of the evidence is zero.

Evidentiary value of data

The results of the several analyses are generally insufficient to reduce the suspect population to one person only. Therefore, the classical discrimination or classification tasks, either speaker verification or identification, are inadequate to interpret the evidence because they lead to a binary decision of identification or nonidentification. Moreover, a probabilistic model (the bayesian model) allows for revision based on new information of a measure of uncertainty about the truth or falsehood of an issue. It allows the scientist to evaluate the evidentiary value of data, for instance the comparison of control recording and audio recording evidence, without making a binary decision on the identity of the speaker.

Speaker Profiling

Speaker profiling is a classification task, performed mostly by phoneticians, where a recording of the voice of a perpetrator is the only lead in a case. The classification specifies the sex of the individual, the age group, dialect and regional accent, peculiarities or defects in the pronunciation of certain speech sounds, sociolect and mannerisms.

Intelligibility Enhancement of Audio Recordings

A recording can be degraded by a variety of different types of noise, each of which may suggest a different enhancement procedure. Enhancement techniques fall into two main categories: canonical filters, such as various types of band pass, and simple comb filters. A second and more effective class of enhancement techniques relies on 'signal-dependent filters'. These filters are microprocessor-based and use digital signal processing techniques, such as adaptive filtering and spectral subtraction (**Fig. 4**).

A careful distinction must be made between sound quality, i.e. attractiveness and ease of listening, and intelligibility, i.e. the degree of understanding of

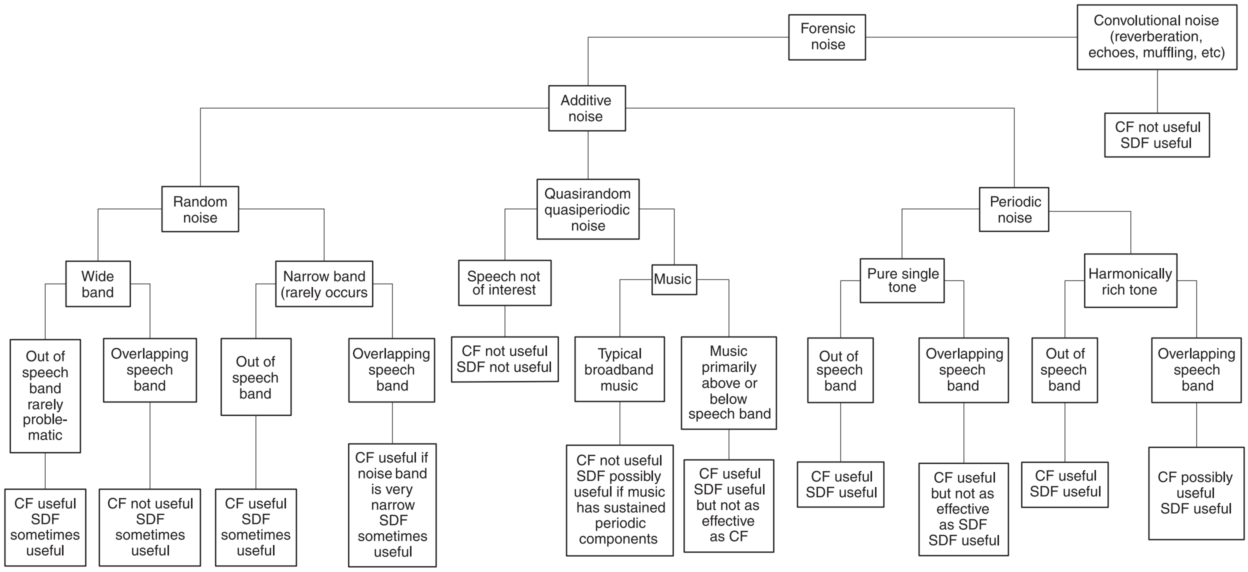


Figure 4 Types of forensic noise and usefulness of enhancement techniques in increasing intelligibility. CF, canonical filtering; SDF, signal-dependent filtering. Copyright © 1993 by Lawyers Cooperative Publishing Company, a division of Thomson Legal Publishing Inc.

spoken language. It is never advisable to decrease intelligibility in order to gain in sound quality. In some situations, background noise or obtrusive noise can be successfully attenuated. However, in many cases, especially where improper recording techniques result in low recorded speech levels or where broad-spectrum random noises are the primary masking sounds, enhancement procedures can be frustratingly ineffective.

Forensic noise is any undesired background sound that interferes with the audio signal of interest, generally speech. It can usually be classified as either additive or convolutional. Additive noise can be attributed to specific noise sources, such as traffic, background music, microphone noise, channel noise, and ambient random noise in cases where the recording level was too low. Convolutional noise, such as reverberations, acoustic resonance and muffling, is the result of the effect of the acoustic environment on the speech sample and on additive noise. Due to the time dispersive and spectral modification effects of these kinds of noise, they can have a more deleterious effect on intelligibility than additive noises.

Transcription and Analysis of Disputed Utterances

The transcription of disputed utterances consists of converting spoken language into written language. If the intelligibility of the speech is optimal, the task can be achieved by a lay person of the same mother tongue as the unknown. If the intelligibility is altered by the speaker or transmission channel distortion, decoding necessitates knowledge of phonetics, linguistics and intelligibility enhancement techniques.

After enhancement or filtering, the sample is first listened to in its entirety to locate particularly difficult passages, learn proper names, and note idiosyncratic characteristics. Proper names are often difficult to decode because there is little or no external information inherently associated with a proper name. If the sample is a conversation, the structure of the dialogue is analyzed in assigning each utterance to a speaker. In sections where the speech is audible but not intelligible, or where there are multiple speakers, the approximate number of inaudible or missing words is counted.

Secondly, the difficult passages are subjected to phonetic and linguistic analysis. The study of linguistic stress and nonlinguistic gestures, especially as highlighted with spectrography, enhances the meaning of the message and knowledge of the word boundaries. Furthermore, they afford suprasegmental information on the sentences and word structure. The analysis of the phoneme placement and manner, as well as the coarticulation, i.e. the fact that each phoneme uttered affects all others near it, provides segmental information on the phonemic composition of words. Once all the possible ambiguities have been resolved, uncertainties have to be clearly mentioned in the final transcript, as the potency of written words is greater than that of spoken words.

Authenticity and Integrity Examination of Audio Recordings

An audio recording used as evidence improves the overall means by which information is communicated to the trier of fact, to the degree that it presents a fair and accurate aural record. However, the ongoing growth of digital technology and the possibility of turning to an audio recording professional create falsifications that are difficult or impossible to detect, and even fabrication of recordings. The debate about the ability to falsify and detect falsifications should be relaunched because it was resolved before digital audio recording and editing became commonplace.

Some types of falsification are inherently more difficult to detect than others. The type of recording medium used, the overall noise level and the type of editing attempted all affect the chances of successfully detecting certain signs or magnetic and electronic signatures. If the recording is originally made in a digital, tapeless format, many of these detection techniques may no longer apply. Thus an audio recording made in a digital, tapeless format will be inherently much more flexible, and thus much more susceptible to alteration, than a similar analog magnetic tape recording.

Certain editing functions are easier to perform than others: the deletion of a syllable is generally much easier to perform than the insertion of that same syllable or a word would be, in part because of coarticulation. Similarly, as the complexity and magnitude of editing increases, the likelihood of its being detected also increases. However, if only a few short, but crucial, deletions or insertions have been made, these would be more difficult to detect and define.

See also: **Evidence**: Statistical Interpretation of Evidence/Bayesian Analysis.

Further Reading

Bolt RH, Cooper FS, David EE, Denes PB, Pickett JM and Stevens KN (1970) Speaker identification by speech spectrograms: a scientists' view of its reliability for legal purposes. *Journal of the Acoustical Society of America* 47:597–612.

Bolt RH, Cooper FS, David EE, Denes PB, Pickett JM and

Stevens KN (1973) Speaker identification by speech spectrograms: some further observations. *Journal of the Acoustical Society of America* **54**:531–534.

Bolt RH, Cooper FS, Green DM *et al.* (1979) *On The Theory And Practice of Voice Identification.* Washington, DC: National Academy of Sciences.

Champod C and Meuwly D (1998) The inference of identity in forensic speaker recognition. *Proceedings of IEEE–ESCA Workshop on Speaker Recognition and its Commercial and Forensic Applications*, Avignon, pp. 125–135.

Clifford BR (1980) Voice identification by human listeners: on earwitness reliability. *Law and Human Behaviour* **4**:373–394.

Cutler PE, Thipgen CR, Young TR and Mueller EB (1972) The evidentiary value of spectrographic voice identification. *Journal of Criminal Law, Criminology and Police Science* **63**:343–355.

Darby JK (1981) *Speech Evaluation in Psychiatry*. London: Grune and Stratton.

Daubert v. *Merell Dow Pharmaceuticals* (1993 US) 125 L Ed 2 469, 113 S Ct 2786.

Furui S (1997) Recent advances in speaker recognition. *First International Conference on Audio- and Video-Based Person Authentification*, Montana (Switzerland), pp. 237–252.

Gruber JS and Poza FT (1995) Voicegram identification evidence. *American Jurisprudence Trials* **54**.

Gruber JS, Poza FT and Pellicano AJ (1993) Audio tape recordings, evidence, experts and technology. *American Jurisprudence Trials* **48**.

Kersta LG (1962) Voiceprint identification. *Nature* **486**:1253–1257.

Koval S, Ilyina O and Khitina M (1998) Practice of usage of auditive and linguistic features for forensic speaker identification. *Proceedings of the 8th COST 250 Workshop on Speaker Identification by Man and by Machine: Directions for Forensic Applications*, Ankara, pp. 23–29.

Koval S, Kaganov A and Kithroy M (1998) The chart of the standard expert actions and decision making principles of forensic speaker identification. *Proceedings of the 8th COST 250 Workshop on Speaker Identification by Man and by Machine: Directions for Forensic Applications*, Ankara, pp. 62–66.

Kunzel HJ (1994) Current approaches to forensic speaker recognition. *Proceedings of ESCA Workshop on Automatic Speaker Recognition Identification and Verification* Martigny, pp. 135–141.

Nolan F (1983) *The Phonetic Bases of Speaker Recognition*. Cambridge: Cambridge University Press.

Reynolds DA (1995) Automatic speaker recognition using gaussian mixture speaker model. *Lincoln Laboratory Journal* **8**:173–191.

Tosi O, Oyer H, Lashbrook W, Pedrey C, Nichol J and Nash EW (1972) Experiment on voice identification. *Journal of Acoustical Society of America* **51**:2030–2043.

W

Weapons *see* **Firearms**: Humane Killing Tools; Laboratory Analysis; Range; Residues; Weapons, Ammunitions and Penetration.

White-Collar Crime *see* **Computer Crime**.

WILDLIFE

Kenneth Goddard and **Edgard Espinoza**, National Fish and Wildlife Forensics Laboratory, Ashland, OR, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0618

Introduction

'Why would the federal government build a separate and fully functional crime laboratory just for wildlife law enforcement?' This is the question most often asked when people hear about the National Fish and Wildlife Forensics Laboratory – the first full-service national or international crime laboratory devoted to wildlife law enforcement – located in Ashland, Oregon. The answer, sad to say, is human greed, coupled with the increasing value placed on wildlife parts and products derived from protected, threatened and endangered species. 'But how much can a poached deer be worth?' these questioners often respond, thinking, as most people do, that wildlife law enforcement mostly involves individual hunters violating local hunting season rules and regulations. Which is to say, mostly misdemeanor violations.

Well, in a comparative sense, a single doe taken out of season probably isn't worth all that much. But most deer species are relatively plentiful at the moment, whereas the world is rapidly running out of elephants, tigers, leopards, jaguars, bears, and literally hundreds of other species. Which is why you can find elephant leather briefcases and elephant ivory carvings on sale for several hundred dollars apiece; tiger claw necklaces and leopard/jaguar coats in the $5000–10 000 range; and 'canned' hunts for grizzly bears starting at $15 000. And the list goes on in a progressively depressing manner.

But all of this should not come as much of a surprise. After all, to a great extent, it is just human nature. People want to buy things that are rare and/or unusual ... either to possess themselves or to give as special gifts. And what could be more rare or usual than a shatoosh shawl made from the microscopically fine hairs of the fleece of the Tibetan antelope (*Pantholops hodgsonii*), or a fine long-stemmed comb made from the shell of a 40-year-old Ridley sea turtle, or a 1 oz (30 g) tin of caviar from the endangered shortnose sturgeon?

And then we get into cultural issues, and discover that many of these wildlife parts and products are considered critical ingredients of a wide array of traditional Asian medicines that have been used for thousands of years, which results in hoarded rhino horns selling for $75 000 each, bear gall bladders for a $1000 a gram, and tiger bones for whatever the market will bear – which is quite a lot these days.

And we cannot forget the underlying philosophical question: whether or not species should simply be allowed to go extinct through the process of 'natural' selection, whatever that may or may not mean.

But, in essence, the wildlife problem really comes down to the simple issue of supply and demand on a

national and international scale. Much like the illicit drug trade, if there is a demand for an illicit wildlife part or product, and a great deal of money to be had, then enterprising criminals will find a way to supply the goods. And you can be sure these suppliers will not be overly concerned about national or international borders, much less the possibility that these endangered and threatened species may well disappear before the next generation of suppliers can seek their illicit profits.

Thus the mission of the National Fish and Wildlife Forensics Laboratory: to provide forensic support for game wardens, conservation officers and wildlife special agents throughout the world. And, in doing so, we often find ourselves involved in felony investigations as well as misdemeanors, because it is a felony to violate the US Endangered Species Act, the penalty for which may be as much as a $100 000 and a year in federal prison per count. And all of a sudden, trafficking in wildlife parts and products becomes serious business for all concerned.

So how does a national and international wildlife crime laboratory go about its assigned mission to provide forensic support to state, national or international wildlife law enforcement officers? Well, the answer to that question is a bit complex. And as you will soon see, it takes us right back into thorny issues of philosophy and culture as well as science.

But first, let us start at the beginning.

Basic elements of wildlife forensic science

As in any police crime laboratory, forensic experts employed by a wildlife crime laboratory are expected to do two things:

1. identify their evidence;
2. in a triangular fashion, link suspect, victim and crime scene together with that physical evidence.

The only significant difference is that, in a wildlife investigation, the victim is almost always an animal, and every now and then it turns out that the suspect is also an animal. Thus, the first obligation of a wildlife forensic scientist is to distinguish the act of one animal fighting and killing another (typically for food, territory or mates) from acts of human greed and avarice against animals. It is only then that wildlife forensic scientists can begin to apply the principles and techniques of forensic science to the investigation of wildlife crimes – the point at which the problems truly begin.

In the following paragraphs, we hope to provide the reader with an understanding of these problems, a history of wildlife forensics, and an overview of the forensic principles and techniques currently being applied to wildlife-related evidence, as well as a sense of how we think wildlife forensics might evolve in the coming years.

Problems of Wildlife Forensics

At a homicide scene, the task of linking the suspect to the victim and the crime scene is often complicated by the difficulty of determining what the scene was like before the suspect and victim came into lethal contact. In a like manner, the difficulty of linking the suspect and victim in an illegal hunting situation is frequently complicated by:

- the circumstances under which an animal can be legally killed;
- the lack of species-specific definitions for wildlife parts and products.

Legal versus illegal kills

In a sense, human forensic scientists have it easy because it is almost always illegal to kill a human being (except in self-defense or in an act of war). However, the circumstances under which an animal can be legally hunted and killed are often quite complex. Examples of this complexity can be found in the following questions frequently posed to a crime laboratory by wildlife investigators:

- *How was this animal hunted and killed?* Legal hunting may be restricted to the use of specific types of weapons (archery, black-powder firearms, etc.) to kill certain animals during a defined hunting season, or within a defined hunting area.
- *When did this animal die?* It is illegal to kill migratory birds before sunrise or after sunset during a defined hunting season, or illegal to kill most animals the day before a defined hunting season.
- *Where was this animal killed?* It may be legal to hunt in a federally protected wildlife refuge during a certain hunting season, but illegal to do so in adjoining private property.
- *How many animals were killed by this hunter during this hunting season?* It may be legal to kill one animal of a specific species during a defined hunting season, but not more than one.
- *What are the genders of the animals killed?* It may be legal to kill a male whitetail deer during a defined hunting season, but not a female.
- *Which hunters were involved in the illegal hunting?* In a typical waterfowl hunting blind situation, where several hunters may be shooting in the same area, and 'dropped' ducks tend to be comingled, it may be difficult to tell which hunters killed 'over their limit' and which hunters hit nothing at all. The issue can be further complicated if not-to-be-

exceeded point values are assigned to certain species of ducks during a hunting day.

- *Was the animal killed in self-defense?* It may be legal for a hunter to kill an otherwise completely protected endangered species in self-defense: if, for example, the bullet trajectory indicates that animal was charging the hunter instead of running away at the time of the shooting.

Lack of species-specific definitions for wildlife parts and products

In a police crime laboratory, the task of identifying evidence tends to be relatively simple. 'The white powder is 86% cocaine hydrochloride.' 'The blood sample contains 0.20% ethyl alcohol.' 'The lethal projectile is a .40 caliber copper-jacketed hollow-point bullet.' In these examples, the task is simple because there are simple and straightforward definitions for these items that are accepted worldwide.

But in a wildlife crime laboratory, the submitted evidence items are usually parts and products from endangered, threatened or protected species of animals. And while underlying definitions for all of these species do exist in the form of morphological or genetic 'keys', the definitions are almost always based upon knowing the country of origin (especially for subspecies) and having the entire animal available for examination. And therein lies the problem. Wildlife investigators rarely seize whole animals as evidence. Instead, they tend to seize parts and products of animals from boats, planes and warehouses. As such, they rarely know the country of origin of these items, and cannot depend on those who do know (usually, the suspects) to provide a dependable answer.

Given this inability to use standard taxonomic keys to determine species of origin, wildlife forensic scientists must conduct extensive research to come up with new species-defining characteristics that will enable them to stand up in a court of law and testify that a part or product in question came from a certain species, and no other species in the world. And it is the 'no other species in the world' part that makes the species-source identification of a wildlife part or product a formidable task indeed, especially when faced with over 4629 species of mammals, 9682 species of birds and 7962 species of reptiles.

Brief History of Wildlife Forensics

Early species identification for forensic purposes relied on visual comparisons, and microscopic 'analysis' of blood components. The advent of immunological techniques allowed an analyst to determine the Family origin of tissues; and, in a like manner, electrophoretic analysis of proteins enabled the analyst to accurately determine species origin. But it was not until the development of restriction fragment length polymorphism (RFLP) and polymerase chain reaction (PCR) techniques, and the analysis of the mitochondrial DNA sequence, that wildlife forensic scientists were able to address subspecies issues, and to compare blood or tissue samples as having come from a single common animal source.

All of this began with the first combined wildlife management and wildlife forensic laboratory established by the State of Wyoming in 1965. Other small and combined wildlife management and forensic state laboratories were created during the next two decades, but it was not until 23 years later that a full-service national wildlife forensic laboratory was created that focused only on forensic issues and did not address any other wildlife conservation issues. In 1988, the National Fish and Wildlife Forensics Laboratory was established for the unique purpose of assisting federal, state and international law enforcement officers in their investigations of wildlife law violations.

Current Science of Wildlife Forensics

While the science of wildlife forensics is still very much in its infancy, numerous protocols have been established for the identification and comparison of wildlife-related evidence. These protocols are typically divided into the following analytical categories:

- pathology;
- molecular biology;
- morphology;
- criminalistics;
- chemistry.

Pathology

Pathology is basically the study of death (**Figs 1** and **2**). And in a wildlife crime laboratory, veterinary pathologists are responsible for determining the cause of death of an animal. They are able to accomplish this by conducting a necropsy (autopsy) that involves a search, not only for lethal wounds caused by bullets, arrows, spears and traps but also a comprehensive toxicological examination of blood, urine, tissue and stomach/crop contents to eliminate or confirm a poison or contaminate cause of death. In doing so, the pathologist also makes a corresponding search for signs of disease vectors that may indicate a natural cause of death.

Issues that may complicate determination of cause of death in an animal, but certainly do not affect the results of a careful and professional necropsy examination, include:

Figure 1 (see color plate 48) Dr Richard Stroud and Elaine Plaisance examine a North American bear cub to determine cause of death.

- the possibility that the animal may have been struck by additional bullets, pellets or other projectiles (that were not lethal or crippling) for a period of months or years;
- the relatively common practice of illicit bow-hunters shooting at an animal with a firearm first (because it is difficult to get close enough to a suspicious animal for a killing show with a bow), and then sticking an arrow into the bullet wound;
- the possibility that the animal was killed or fatally weakened by a modern pesticide or poison that was purposefully designed to decompose rapidly after a few hours of exposure to air or sunlight;
- the likelihood that scavengers will rapidly destroy useful blood, tissue or bone evidence if the carcass is not found and collected in a matter or hours or a very few days.

In conducting necropsies involving bullet wounds, it is often extremely important that the veterinary pathologist determine the trajectory of the bullet into or through the body. This information may resolve the question of whether the accused hunter was properly defending himself or herself against a charging animal, or illegally hunting a protected species.

Molecular biology

Molecular biology is, among other things, the study of genetic information encoded in the DNA molecule and the expression of that coding into proteins and related biological structures (**Fig. 3**). Given the wide variety of biological structures present in the animal kingdom, molecular biology offers the wildlife forensic scientist an extremely powerful tool to:

- determine family, genus and species;
- determine gender;
- individualize blood and tissue samples.

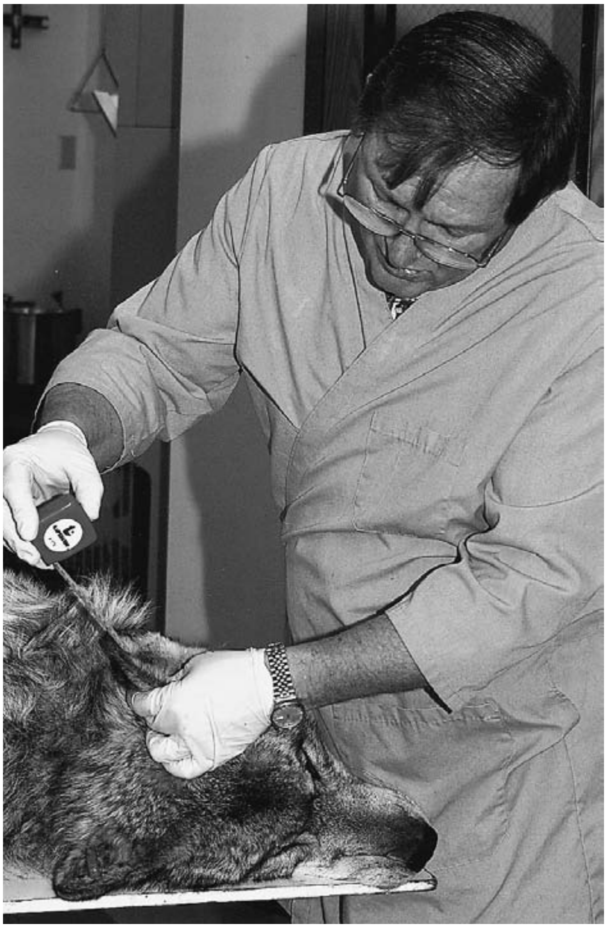


Figure 2 Dr Richard Stroud conducts a necropsy on a North American wolf.

Figure 3 Caviar, the oocytes of sturgeon, is analyzed to determine species of origin. Analysis is done by sequencing part of the mitochondrial DNA of a single egg and each sequence is compared against reference standards worldwide of sturgeon species.

Family/genus/species identification The process of determining the species origin of an unknown tissue normally begins with a series of immunological tests designed to narrow the possibilities down to the species comprising a single family (e.g. bears – Family Ursidae, or deer – Family Cervidae). At that point, the examiners can go forward with either protein or DNA/PCR analysis (along with related comprehensive databases) to determine the actual species involved.

Gender identification There are a number of nuclear DNA-based gender-determining tests available for mammalian species. They generally use PCR amplification to detect specific sequences of the *ZFY* and/or *SRY* genes, both of which are located on the mammalian Y chromosome.

Individualization of blood and tissue Initial work on individualizing animal blood and tissue samples involved multilocus DNA probe hybridization techniques. However, new PCR methods for detecting single-locus single tandem repeat (STR) markers have been developed and applied in human forensic casework, demonstrating the technical feasibility of similar applications to animal species. Several wildlife forensic laboratories are currently developing STR markers for determining the individual origins of wildlife evidence tissues.

Morphology

Morphology is the study of structure. In wildlife forensic science, morphological examinations of submitted evidence items are typically conducted by eye or with the use of simple, compound or scanning electron microscopes. They are often the simplest examinations performed in a wildlife crime laboratory; but at the same time, they represent some of the most complex and difficult identification problems (**Figs 4–7**).

As an example, the authors of this article are both absolutely convinced that they could positively identify their daughters from 10 000 similar young women. But could they create a written protocol that would enable another forensic scientist to make that same positive identification with the same degree of certainty? Very likely not. And the reasons why such a protocol would be a difficult to write lie in the heart of the morphological problem: the lack of standard definitions.

Let us go back to that cocaine hydrochloride molecule for a moment. An underlying tenant of organic chemistry is the simple fact that one cocaine hydrochloride molecule looks exactly like every other cocaine hydrochloride molecule in the world. So, in a relative sense, it is easy for crime laboratories throughout the world to come up with a single protocol to identify cocaine hydrochloride. But what happens when a crime laboratory shifts its focus to whitetail deer?

First of all, you discover – outward appearances aside – that whitetail deer are not alike. In fact, no two whitetail deer are exactly alike (even if they are genetic twins!) because the normal wear and tear that a young whitetail deer experiences, literally from the moment of birth, create individual characteristics (a healed cut, a chipped hoof, a broken antler) that invariably show up in the animal's morphology.

Figure 4 Bonnie Yates examines the remains of a wolf carcass that has been cleaned in the laboratory by dermestid beetles. These beetles are an essential part of cleaning mineralized tissue so that the species-defining osteological components are clearly visible.

Figure 5 Bonnie Yates and Cookie Sims examine an evidence item against a bear paw morphological standard. Bear paws are found in trade for the use of the claws in jewelry and in certain Asian medicinal preparations.

Figure 6 Dr Pepper Trail and Bonnie Yates examine osteological remains of endangered and protected birds recovered from a crime scene.

Figure 7 Dr Edgard Espinoza examines an ivory statue. Trade in ivory from African and Asian elephants is prohibited, whereas trade in the ivory from the extinct mammoth is legal.

So the immediate problem for a wildlife forensic scientist is to come up with *class characteristics* that distinguish and identify family, genus and species of animals that are separate and distinct from population and individual characteristics.

These identification characteristics typically fall within one of the following morphological categories:

- hair and fur;
- leather and hides;
- bones and skulls;
- teeth, claws and beaks;
- hooves, horns and antlers;
- feathers and down;
- other miscellaneous parts.

Criminalistics

The ability of a wildlife forensic scientist to process evidence from an illegal hunt, much like that of a homicide scene, often comes as an unpleasant surprise to the unwary poacher. In fact, the events associated with a typical illegal hunt often involve the following categories of 'criminalistics' (or police forensic science) evidence:

- trace evidence;
- firearms;
- other weapons;
- impression marks;
- latent prints;
- questioned documents.

Trace evidence Trace evidence in an animal case can involve a wide range of materials. A classic example is a case in which a mountain lion was held captive for a period of time and then killed in an illegal 'canned' hunt. As it turned out, the mountain lion tried to chew his way loose from the synthetic fiber ropes (two types were used by the suspects to secure the lion), and a forensic scientist was able to link the fibers from the lion's stomach back to chewed ropes found at the crime scene.

Firearms The typical circumstances in which an animal is killed with a firearm vary greatly from those of a homicide case. The most significant differences include:

- the distance from suspect to victim;
- the choice of firearm;
- the ability of the suspect to 'clean up' the scene;
- the tendency of the suspect to take the victim from the scene;
- the tendency of the suspect frequently to reuse the same firearm.

Unlike human crime situations, in which the victim is most commonly killed with a pistol at short – 'contact' to 25 yard (23 m) – distances, the typical animal kill involves a high-powered (and large-caliber) rifle or a shotgun at relatively long distances (50–300 yards (45–275 m)).

Given the nature of the typical hunting area (brush, trees and ground cover), it is often difficult for an illicit hunter to retrieve expended casings; however, the long shooting distances and the fact that the shot could have come from any 360° vector point makes it extremely difficult for a crime scene investigator to locate the shooting point, much less the expended casings. But all of these advantages (to the illicit hunter) tend to be negated by two simple facts:

1. The whole point of the illicit hunting is for the suspect to take the victim (as a trophy or meat) back to his or her residence. Thus, the bullet is likely to either be in the carcass of the animal, or in the 'gut pile' left at the scene (which can be matched to the trophy head or meat with DNA techniques).

2. The typical illicit hunter spends a lot of money on a rifle or shotgun, and is rarely willing to discard the weapon after a single illicit kill. Thus, it is very likely that a succession of illegal kills can be linked to a single poacher by matching the spent bullets or casings to the rifle or shotgun.

Other 'weapons' Other hunting weapons typically associated with an illicit animal kill include:

- long bows and arrows;
- crossbows and 'bolts';
- spears;
- spring traps;
- poison discharge devices;
- nets.

Impression marks The fact that most illicit hunting situations occur in remote areas or 'off road' situations makes it extremely likely that the suspect will leave tire tracks and boot impressions in soil, mud or snow. And the fact that these tires and boots are typically used in off-road situations makes it all the more likely that the tire or boot treads will possess individualistic wear marks.

Latent fingerprints Latent fingerprints are the classical means of linking suspected, victim and crime scene through physical evidence (**Fig. 8**). The following types of latent-bearing evidence are frequently submitted to a crime laboratory in wildlife cases:

- firearms;
- expended casings (shotgun and rifle);
- knives;
- 'No Trespassing' and 'No Hunting' signs;
- game tags;
- import/export (CITES) permits.

Questioned documents A frequently-asked question in a national or international wildlife investigation is whether or not the paperwork (typically import/export documents) is valid. This can be an extremely difficult question to answer when the authorizing seals vary between countries, the names of individuals authorized to approve import/export documents change on a frequent basis, and the shipments must be 'cleared' (the documents examined and approved) at the local port of entry.

Questioned documents typically associated with a wildlife case include:

- forged or altered hunting licenses;
- forged or altered game tags;
- forged or altered import/export permits.

Figure 8 Andy Reinholz compares friction ridge evidence lifted from a container found at the crime scene against standard prints collected from a suspect.

Chemistry

Chemical analysis techniques are most often used in a wildlife crime laboratory to provide toxicological information to a veterinary pathologist. A forensic scientist assigned to this section of the laboratory therefore spends a great deal of time examining blood, urine, tissue and stomach/crop contents in a search for pesticides and poisons. However, analytical chemistry procedures (that utilize chemical biomarkers) are also used to identify the species source of animal products, such as bear bile and deer musk, and standard toxicological methods are routinely employed to identify chemical baits and poisons used to trap and kill wildlife (**Figs 9** and **10**).

Future of Wildlife Forensics

Where is all of this taking us? Into some fascinating areas of technology, no doubt, where rapid DNA sequencing techniques may enable us to crack the genetic code that distinguishes family, genus, species

Figure 9 An examiner prepares the bile salts from an importer's seized medicinal products to determine if the samples are of ursid provenance. The bile acids extracted from a bear's gallbladder are used in some traditional Asian medicinal preparations.

Figure 10 Dr Mark Kirms examines blood stains by electrospray mass spectrometry to determine species identities from hemoglobin molecular weights.

and subspecies. And very likely into an era where game wardens and conservation officers routinely collect blood and tissue samples from the sites of illegal kills, knowing full well that the resulting genetic data will be placed in the databases of wildlife forensic laboratories – awaiting the submission of matching evidence. Thus, poaching scenes will be routinely worked in the manner of a modern homicide investigation involving a human victim.

But regardless of where newly developed technologies take us, we will almost certainly still find ourselves facing those same thorny issues of culture and philosophy that make the practice of wildlife forensics a fascinating and thoroughly satisfying career choice.

See also: **Administration of Forensic Science**: An International Perspective. **Criminalistics**. **Fingerprints (Dactyloscopy)**: Visualization. **Firearms**: Weapons, Ammunitions and Penetration.

Further Reading

Aasen E and Medrano JF (1990) Amplification of the ZFY and ZFX genes for sex determination in humans, cattle, sheep and goats. *Biotechnology* 8:1279–1281.

Adamczyk M and Gebler JC (1997) Electropspray mass spectrometry of a and b chains of selected hemoglobins and their TNBA and TNB conjugates. *Bioconjugate Chemistry* 8:400–406.

Andrasko J and Rosen B (1994) Sensitive identification of hemoglobin in bloodstains from different species by high performance liquid chromatography with combined UV and fluorescence detection. *Journal of Forensic Sciences* **39**(6):1018–1025.

Espinoza EO, Kirms MA and Filipek MS (1996) Identification and quantitation of source from hemoglobin of blood and blood mixtures by high performance liquid chromatography. *Journal of Forensic Sciences* **41**(5):804–811.

Espinoza EO, Lindly NC, Gordon KM, Ekhoff JA and Kirms MA (1998) Species determination from hemoglobin by electrospray mass spectrometry. (Submitted).

Fain SR and LeMay JP (1995) Gender identification of humans and mammalian wildlife species from PCR amplified sex-linked genes: *Proceedings of the 47th Conference of American Academy of Forensic Sciences*, vol. 1, p. 34.

Irwin DM, Kocher TD and Wilson AC (1991) Evolution of the cytochrome. *Journal of Molecular Evolution* **32**:128–144.

Moore TD, Spence LE and Dugnolle CE (1974) *Identification of the dorsal guard hairs of some mammals of Wyoming*. Bulletin No. 14. Laramie: Wyoming Game and Fish Department.

Oates DW, Brown CW and Weigel DL (1974) *Blood and tissue identification of selected birds and mammals*. JPR Study XII, Project W-38-R. Lincoln, NE: Nebraska Game and Parks Commission.

Sensabaugh GF (1976) Molecular evolution and the immunological determination of species. *International Microform Journal of Legal Medicine* **11**(2):219.

Taylor AJ, Linforth R, Weir O, Hutton T and Green B (1993) Potential of electrospray mass spectrometry for meat pigment identification. *Meat Science* **33**:75–83.

WOOD ANALYSIS

S Palenik, Microtrace, Elgin, IL, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0619

Introduction

Wood is encountered, in one form or another, in connection with a variety of crimes. Its use as evidence is limited primarily to three areas: (1) physical matches between broken pieces of wood; (2) as a substrate carrying toolmarks; and (3) as small pieces or particles of trace evidence. Other questions may be raised in special circumstances. In the first two categories the wood is only incidental. Toolmark examiners routinely perform physical matches and toolmark comparisons on a variety of materials. Although wood may present some special challenges, because of its fibrous nature, a knowledge of wood structure and anatomy is usually unnecessary to perform a successful comparison in these cases. Experience in interpreting physical matches and toolmarks and in recording the results are the primary knowledge needed for these types of examination. However, a background in wood technology is useful, as illustrated by the investigations by Arthur Koehler into the ladder used in the kidnapping of the Lindbergh baby. Koehler was able to trace pieces of the handmade wodden ladder back, not only to the mills that sawed the wood, but to the lumberyards which sold some of the boards, based on his intimate knowledge of wood and the lumber industry in all their aspects. After Richard Hauptman was identified as a suspect, Koehler was able to locate a board in his attic which compared in all respects to one of the rails from the ladder. Typical forensic wood examinations are not this elaborate. Most involve the identification of small pieces of wood or sawdust, and normally the only comparison that can be made is to complete accurate identifications of both the questioned and known wood samples. Exceptions occur when the wood is painted, impregnated or coated with something which permits further comparisons to be conducted. In a recent homicide investigation, detectives discovered that the murderers had broken down the door of the apartment in which the victim lived, showering down wood particles and plaster as the door gave way. Vacuum sweepings from the shoulders of the suspects' clothing revealed the presence of the same genera of wood comprising the door frame and molding, as well as small particles of calcium sulfate consistent with plaster. One of the suspects was employed in millwork, but none of them had an explanation for both the wood and plaster.

Andrasko J and Rosen B (1994) Sensitive identification of hemoglobin in bloodstains from different species by high performance liquid chromatography with combined UV and fluorescence detection. *Journal of Forensic Sciences* **39**(6):1018–1025.

Espinoza EO, Kirms MA and Filipek MS (1996) Identification and quantitation of source from hemoglobin of blood and blood mixtures by high performance liquid chromatography. *Journal of Forensic Sciences* **41**(5):804–811.

Espinoza EO, Lindly NC, Gordon KM, Ekhoff JA and Kirms MA (1998) Species determination from hemoglobin by electrospray mass spectrometry. (Submitted).

Fain SR and LeMay JP (1995) Gender identification of humans and mammalian wildlife species from PCR amplified sex-linked genes: *Proceedings of the 47th Conference of American Academy of Forensic Sciences*, vol. 1, p. 34.

Irwin DM, Kocher TD and Wilson AC (1991) Evolution of the cytochrome. *Journal of Molecular Evolution* **32**:128–144.

Moore TD, Spence LE and Dugnolle CE (1974) *Identification of the dorsal guard hairs of some mammals of Wyoming*. Bulletin No. 14. Laramie: Wyoming Game and Fish Department.

Oates DW, Brown CW and Weigel DL (1974) *Blood and tissue identification of selected birds and mammals*. JPR Study XII, Project W-38-R. Lincoln, NE: Nebraska Game and Parks Commission.

Sensabaugh GF (1976) Molecular evolution and the immunological determination of species. *International Microform Journal of Legal Medicine* **11**(2):219.

Taylor AJ, Linforth R, Weir O, Hutton T and Green B (1993) Potential of electrospray mass spectrometry for meat pigment identification. *Meat Science* **33**:75–83.

WOOD ANALYSIS

S Palenik, Microtrace, Elgin, IL, USA

Copyright © 2000 Academic Press

doi:10.1006/rwfs.2000.0619

Introduction

Wood is encountered, in one form or another, in connection with a variety of crimes. Its use as evidence is limited primarily to three areas: (1) physical matches between broken pieces of wood; (2) as a substrate carrying toolmarks; and (3) as small pieces or particles of trace evidence. Other questions may be raised in special circumstances. In the first two categories the wood is only incidental. Toolmark examiners routinely perform physical matches and toolmark comparisons on a variety of materials. Although wood may present some special challenges, because of its fibrous nature, a knowledge of wood structure and anatomy is usually unnecessary to perform a successful comparison in these cases. Experience in interpreting physical matches and toolmarks and in recording the results are the primary knowledge needed for these types of examination. However, a background in wood technology is useful, as illustrated by the investigations by Arthur Koehler into the ladder used in the kidnapping of the Lindbergh baby. Koehler was able to trace pieces of the handmade wodden ladder back, not only to the mills that sawed the wood, but to the lumberyards which sold some of the boards, based on his intimate knowledge of wood and the lumber industry in all their aspects. After Richard Hauptman was identified as a suspect, Koehler was able to locate a board in his attic which compared in all respects to one of the rails from the ladder. Typical forensic wood examinations are not this elaborate. Most involve the identification of small pieces of wood or sawdust, and normally the only comparison that can be made is to complete accurate identifications of both the questioned and known wood samples. Exceptions occur when the wood is painted, impregnated or coated with something which permits further comparisons to be conducted. In a recent homicide investigation, detectives discovered that the murderers had broken down the door of the apartment in which the victim lived, showering down wood particles and plaster as the door gave way. Vacuum sweepings from the shoulders of the suspects' clothing revealed the presence of the same genera of wood comprising the door frame and molding, as well as small particles of calcium sulfate consistent with plaster. One of the suspects was employed in millwork, but none of them had an explanation for both the wood and plaster.

Wood Identification

Sample preparation

The identification of wood requires a knowledge of wood anatomy as well as skill in sectioning and mounting small particles for microscopical study. Particles of wood may be matchstick sized or smaller; all the way down to sawdust and sometimes even single fibers. The microscopical identification of wood from small particles requires that the specimen be sectioned in three specific directions to reveal the characteristic morphological features necessary for identification. A piece of wood, no matter how small, retains its orientation with respect to the tree in which it originally grew. Thus a *transverse* or cross-section lies in a plane perpendicular to the length of the tree. It is the plane one observes on the stump of a tree which has been sawn down. On a transverse section one can observe lines of cells radiating from the center of the tree, called rays, which are oriented like the spokes of a bicycle wheel. A section cut in a plane parallel to the length of the tree and aligned with one of these rays is called the *radial* section. The third direction lies in a plane parallel to the length of the tree, which cuts through the rays at a right angle and is referred to as the *tangential* section. These sections are therefore tangential to the circumference of the tree from which the wood originated. Thus, transverse sections are oriented perpendicular to the elongated fibrous cells comprising the wood, and radial and tangential sections are oriented parallel to the length of these fibers. These directions are not difficult to locate on a tree or a large piece of wood, but become a challenge as the size of the specimen decreases.

The first step in preparing a small sample of wood for identification is orientation. Since most pieces of wood, with diameters of the order of a toothpick, are elongated parallel to the fibrous elements of which they are composed, the easiest way to determine orientation is to cut a cross-section straight through the sliver using a sharp razor. The freshly cut surface is then held in the fingers (or if it is too small in a pair of forceps) and examined under a stereomicroscope. The directions of the rays will immediately reveal the true orientation of the fragment. The specimen is now oriented so that the required sections can be cut from it. If the piece of wood is large enough, it can either be mounted for sectioning in a microtome, or sections can be cut freehand with a microtome knife blade using a technique similar to that employed in peeling an apple. In most cases, however, the particles are much smaller and must be sectioned while being observed through a stereomicroscope. If a microtome with a freezing stage is available, the oriented particle may be placed on the stage of a sledge microtome, frozen in a drop of water in which a little gum is dissolved and sections sliced off. Sections may also be cut on a conventional rotary microtome after being impregnated and embedded in wax. This method requires considerable time in sample preparation. For most purposes, sections satisfactory for microscopical identification can be prepared under a stereomicroscope using a sharp single-edged razor. The oriented fragment is held down with an index finger or in a pair of forceps and a small slice is made along the proper direction. The remainder of the section is then peeled away with a sharp pair of pointed forceps. This has the effect of producing sections which are very thin on at least one edge, meaning that almost every section will be a good one as long as it was cut in the right direction. Once the sections have been cut they are placed on a microscope slide and mounted beneath a coverslip in a drop of a mixture of equal parts (v/v) glycerin and alcohol. They are then boiled to remove the air. **Fig. 1** shows a freehand section prepared from a casework-sized sample. Particles of sawdust, which are too small to section, are sprinkled on a slide so that they do not lie on top of one another, boiled in glycerin-alcohol and observed directly under the microscope. Many of the sawdust particles will be cut in a generally radial or tangential section and, therefore, will often show enough characteristics to permit an identification to be made.

When comparing particles of wood in casework, it is essential that good known samples be obtained. A broken door and frame may produce particles from several different kinds of wood. If adequate known samples are not taken at the crime scene, the microscopist may erroneously eliminate questioned wood particles found on a sledge hammer suspected in a break-in, for instance, from being associated with the incident. This problem can become quite taxing in

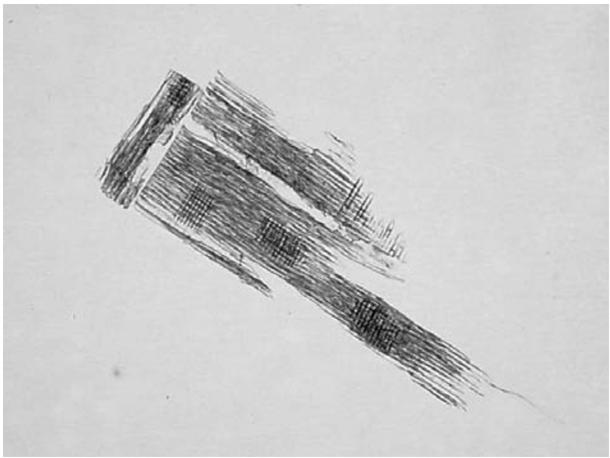


Figure 1 Freehand radial sections from wood fragment after boiling in glycerin alcohol. Original magnification × 40.

certain situations, such as when sawdust from a workshop or particle board are involved. The number of different woods which may be involved can mean that many samples may have to be prepared and examined over a relatively long period of time before a conclusion can be reached.

Wood anatomy

The microscopic features by which wood is identified have been studied and compiled over almost two centuries. Wood is the highly lignified, fibrous tissue (xylem) which is differentiated inside the cambium of a stem towards its center. As a wood-bearing tree grows from a stem, the xylem increases in proportion until it is the dominant tissue in the plant. Although it is possible to name a complete tree in great detail, giving not only its genus and species but perhaps a varietal name as well, this is normally not the case with the wood alone. In most cases, wood can be identified only to genus. On occasion, identification may be carried down to the species level by an experienced microscopist well trained in wood anatomy. If the wood is not exotic, a microscopist trained in wood identification will recognize the common woods almost at sight, based on a few key characteristics. An exotic wood which is rare in commerce will normally make the best evidence. Particles of wood recovered from the scalp of a murder victim were identified as a species of *Kokka*, a rather rare tropical wood not commonly found on the market. An examination of a pool which belonged to the prime suspect showed that it was made from the same wood and had sustained damage to its thick end. This was the most important piece of physical evidence in the trial which resulted in the conviction of the defendant.

The first step in a wood identification based on microscopic anatomy is to decide if the sample is a hardwood or softwood. The terms are perhaps unfortunate, as some hardwoods are softer than some softwoods (e.g. hard pines, which are harder than basswood). A more accurate terminology is based on the fact that the softwoods are all gymnosperms and the hardwoods all originate from angiosperms. Transverse sections of a softwood (*Pinus resinosa*) and hardwood (*Acer saccharum*) are shown in **Figs 2** and **3**, respectively. Softwoods are most easily distinguished from hardwoods on the basis of bordered pits which occur on the radial side of the fibrous elements (called tracheids) that make up the gymnosperm xylem. **Fig. 4** illustrates tracheids with their bordered pits in a radial section of *P. resinosa*. It is often possible to make an identification from the transverse section of a hardwood. **Fig. 5** illustrates a transverse section of wood from *Quercus alba* (white oak). The arrangements of the large and small vessels and the single (uniseriate) rows of rays and the broad (multiseriate) rays are quite characteristic of the oaks. **Fig. 6** shows the tangential section of oak. The uniseriate and multiseriate rays can be seen here as they appear 'head on'. On occasion, the arrangement of the vessels in transverse section is so characteristic that identification to the species level is possible For example,

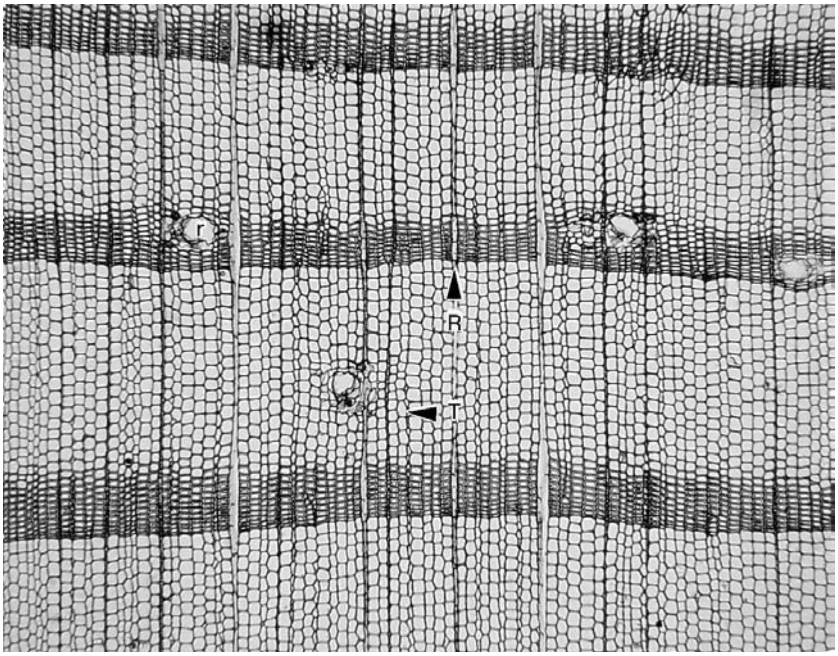


Figure 2 Transverse section of softwood (*Pinus resinosa*). R, radial direction, and T, tangential direction as seen in this orientation; r, resin duct. Original magnification × 40.

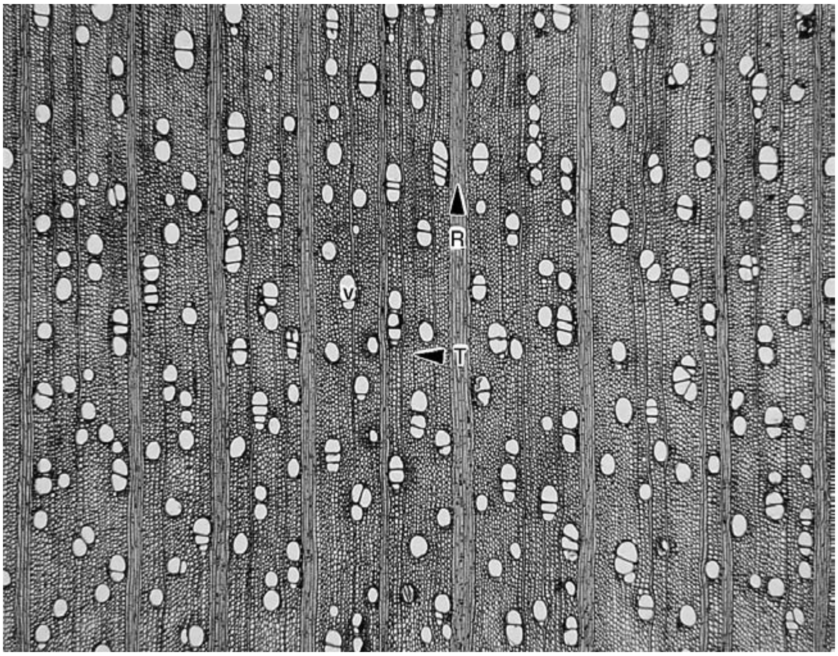


Figure 3 Transverse section of a hardwood (*Acer saccharum*). R, radial direction; T, tangential direction; v, vessel element. Original magnification × 40.

Figs 7 and **8** show the endgrain (transverse section) of *Ulmus fulva* and *U. americana*, two species of elm. The relative sizes, distribution and patterns formed by the large and small vessels make it possible to tell one species from another on the basis of this feature alone. In practice, however, it is often difficult or impossible to make identifications from the transverse sections alone, as the total area of the cross-section is so small that only portions of a few rays and one or two vessels are present. In these cases, radial and tangential sections must be prepared in order to observe other morphological details. **Fig. 9** shows a radial section through *U. fulva* and illustrates the appearance of the vessels and rays in this orientation.

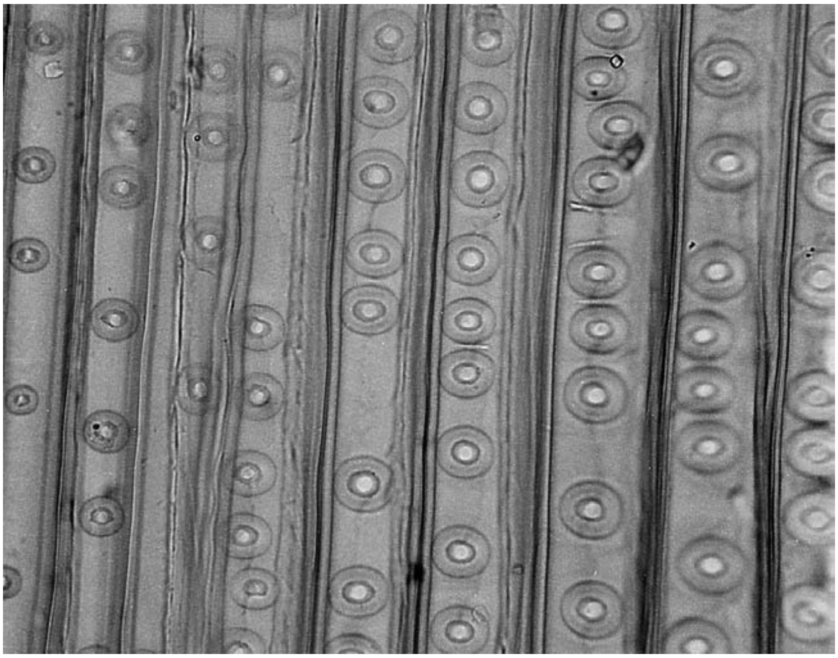


Figure 4 Radial section of wood from *Pinus resinosa* showing vertical tracheids with bordered pits arranged on their surfaces. Original magnification × 383.

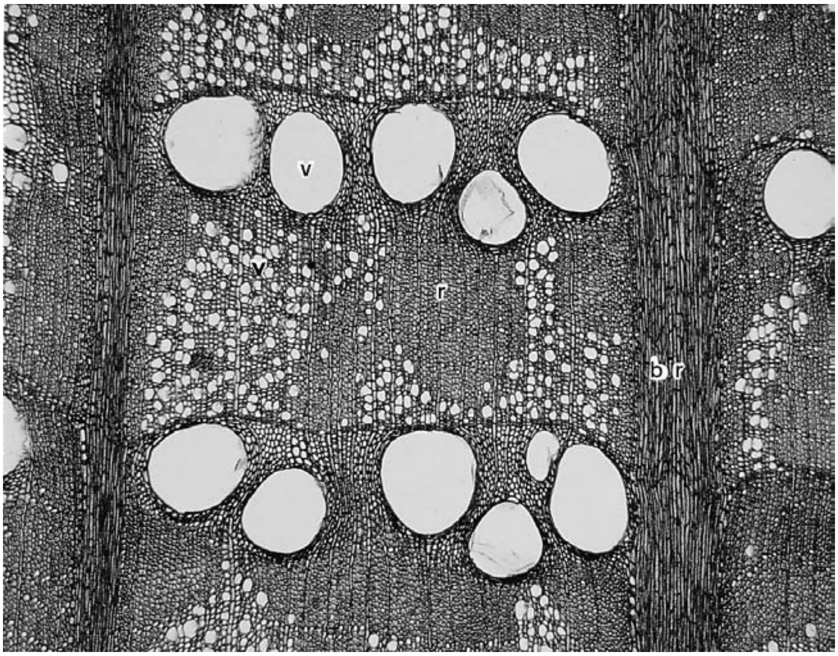


Figure 5 Transverse section of oak (*Quercus alba*). br, multiseriate ray; r, uniseriate ray; v, large and small vessels. Compare with Fig. 5. Original magnification × 40.

Most of the diagnostic structure of softwoods is best observed on the radial section, although transverse sections are still essential for orienting small pieces before sectioning. Resin ducts, which are an important diagnostic character, can best be observed in transverse and radial sections, although they can occasionally be observed in radial sections, as illustrated in **Fig. 10**, which also shows the bordered pits and ray parenchyma cells which make up the rays. **Fig. 11** shows a crossfield region from *P. resinosa* at higher magnification. The four horizontal rows of ray parenchyma cells in the center of the figure contain large window-like pits which are typical of the commercial soft pines and red pine (*P. resinosa*).

Figure 6 Tangential section of oak. br, multiseriate ray; r, uniseriate ray; v, the hole left by a vessel which was torn away in sectioning. Vessels are tubes, composed of cells connected end to end (vessel elements), which are aligned roughly parallel to the other fibrous elements in the wood. Compare with Fig. 5. Original magnification × 40.

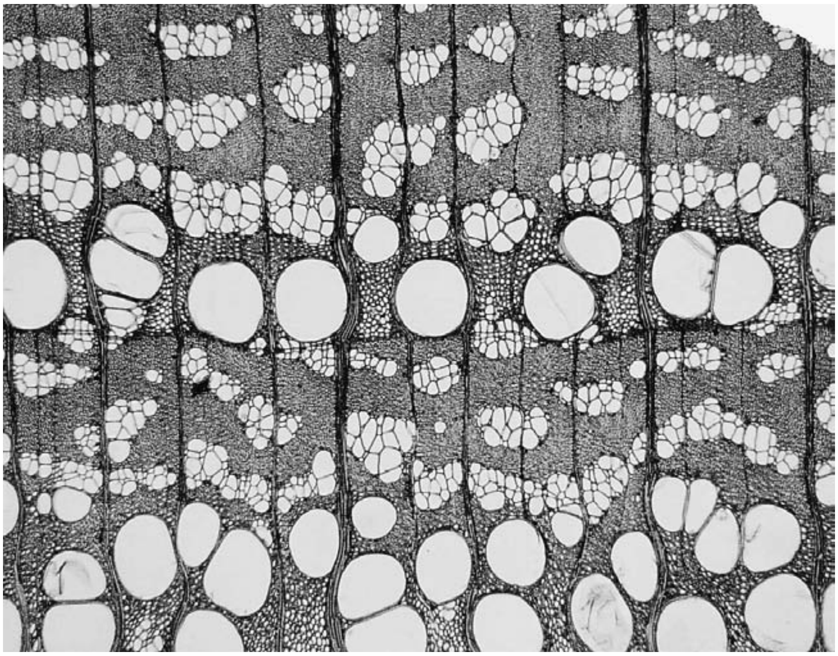


Figure 7 Transverse section of elm (*Ulmus fulva*) showing pattern formed by the large and small vessels. Compare with Fig. 8. Original magnification × 40.

This type of pitting is referred to as 'fenstriform'. The small ray cells with bordered pits above and below these are called 'ray parenchyma'. If these cells contain tooth-like projections, such as those shown in the figure, they are referred to as 'dentate'. The crossfield pitting of spruce (*Picea sitchensis*) is shown for comparison in **Fig. 12.** Note that the single row of ray tracheids above and below the crossfield pits in this wood are not dentate. The small pits in the ray parenchyma cells between the ray tracheids are small and slit-like. These are called 'piceoid pits' and are found also in *Larix* spp. and *Pseudotsuga* spp. Other types of crossfield pits are found in different softwoods.

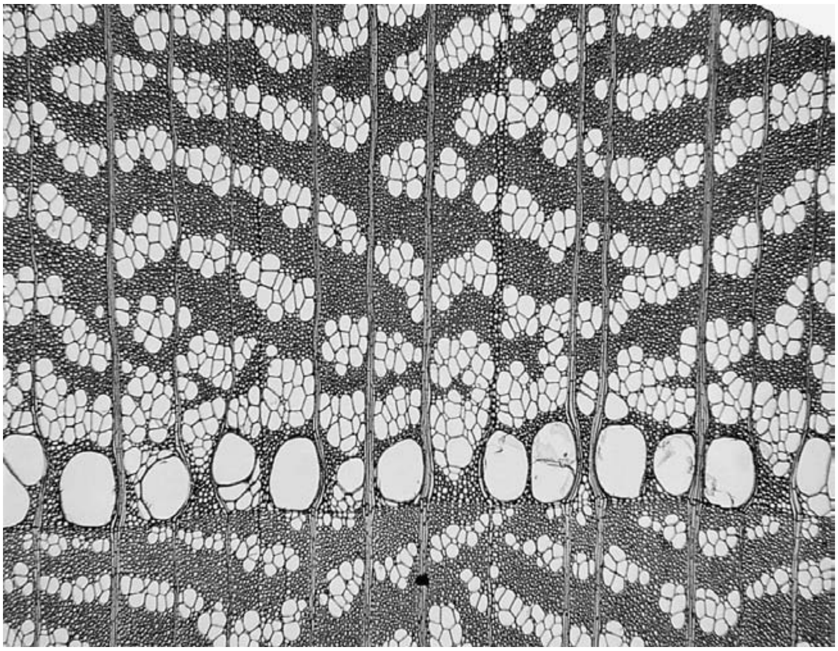


Figure 8 Transverse section of American elm (*Ulmus americana*) showing pattern formed by the large and small vessels. Compare with Fig. 7. Original magnification × 40.

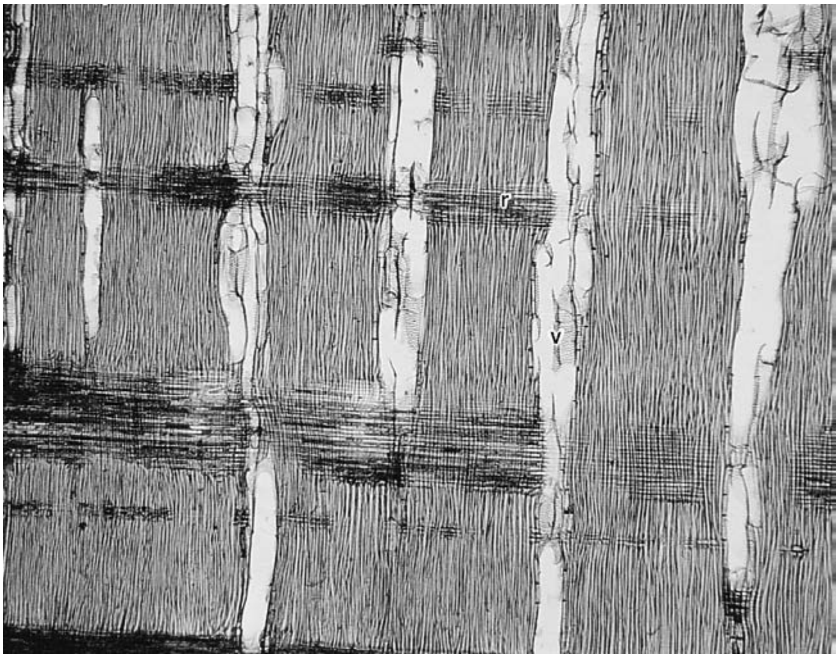


Figure 9 Radial section of hardwood *Ulmus fulva* showing 'brick wall-like' appearance of the rays (r) and pieces of vessel elements (v) which were not torn out of the xylem during sectioning. Original magnification × 40.

Identification

By locating and recognizing the types of features described above, the wood anatomist can narrow the possibilities until only one or a few possible woods remain. For example, the identification of a single cell such as a tracheid bearing bordered pits or a vessel element will serve to distinguish a softwood from a hardwood. Although many common woods can be recognized almost at sight, after experience is gained, small fragments of wood may not show all of the features one would like to see in order to make an identification. Secondly, a single person can only remember so many woods from memory alone.

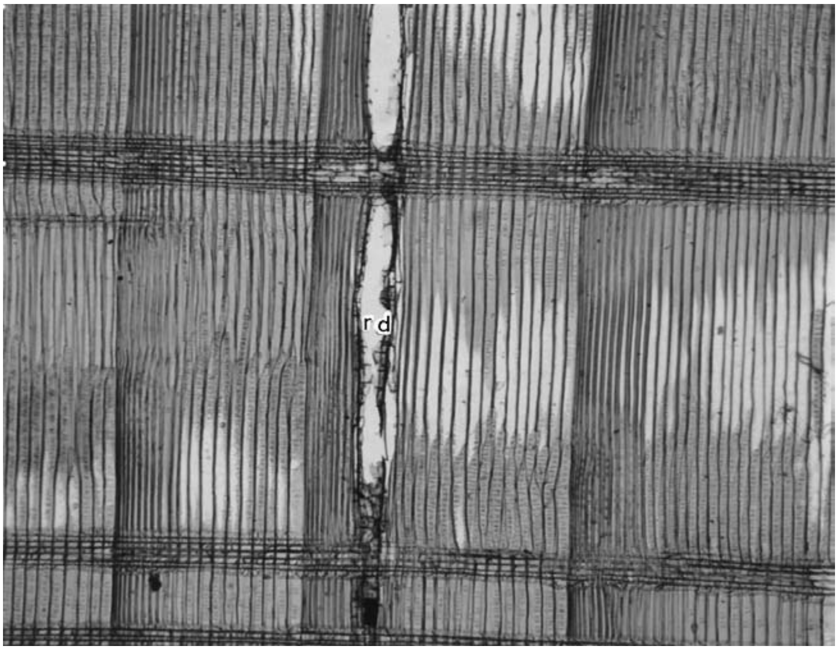


Figure 10 Radial section of softwood *Prunus resinosa* showing 'brick wall' of rays. rd, resin duct. Resin ducts ere essentially empty tunnels, surrounded by a layer of epithelial cells unlike vessels which consist of tubular elements. Original magnification × 40.

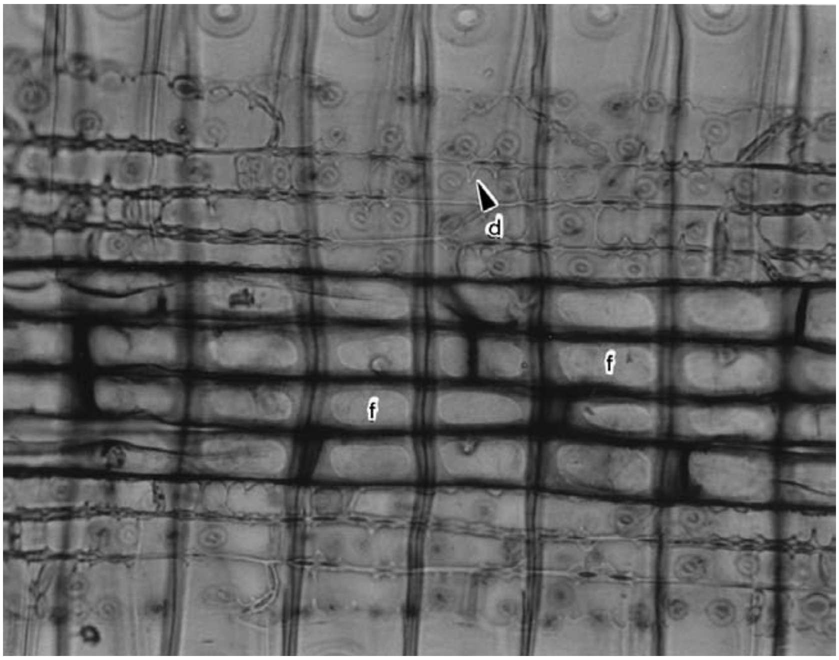


Figure 11 Detail of radial section of *Prunus resinosa* showing horizontal ray parenchyma cells with large window-like pits (f) where they cross over vertical tracheids. Above and below these are horizontal rows of ray tracheids with small bordered pits. The ray tracheids in this wood contain tooth-like projections (d). Cells of this type are called dentate ray tracheids. Original magnification $\times$ 383.

Since the absence of a particular diagnostic feature does not mean that it would not be present if a larger piece of wood were available, the use of a conventional botanical key can be frustrating for identifying woods under these circumstances.

The first solution to this problem was the use of punch cards, which contained coded information on the microscopical anatomical characteristics of hard-and softwoods. The cards were punched or notched around the edges, depending on whether or not a particular characteristic was present in that wood. Each card carried all of the characteristics

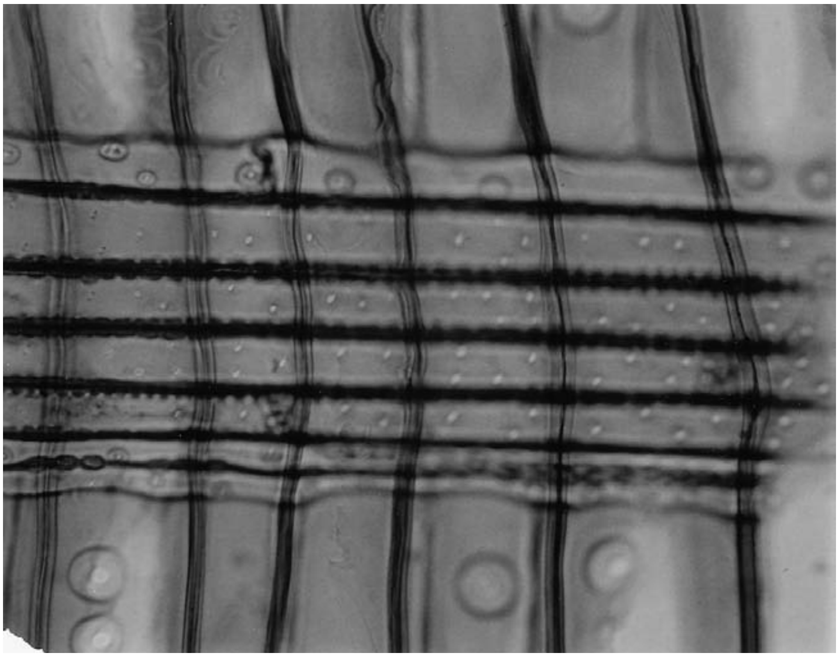


Figure 12 Detail of radial section of spruce wood (*Picea sitchensis*) showing slit-like piceoid pits in the crossfield. The ray tracheids are nondentate. Original magnification $\times$ 383.

exhibited by a particular species. A knitting needle was inserted into the stack and all the cards for woods which did not bear this characteristic would fall out. As the needle was pushed through the stack, the number of cards remaining would become fewer and fewer until only one or a few cards remained. Today these cards have been largely replaced by computer programs, which greatly enhance the microscopist's ability to make identifications from very small pieces of unfamiliar or unusual woods. The International Association of Wood Anatomists has published a glossary of standard terminology for use with these wood identification programs. Once a tentative identification has been made, the specimen must be prepared with known slides or photomicrographs made from authentic wood.

Interpretation of Wood Evidence

Once the particles of wood have been identified, it becomes necessary to interpret the results. The first consideration, aside from the integrity of the evidence itself, must be the accuracy of the identifications. These must be based on a sound knowledge of wood anatomy and comparison to reference specimens of authentic woods, either as physical specimens, prepared microscope slides or photographic atlases. Whenever possible, identifications should be checked by a second analyst, also skilled in the techniques of wood identification. This is a general rule in forensic science and helps to insure the quality of the results obtained. It is also important to take photomicrographs or make drawings of important identification features so that the analyst's notes will reflect the basis for his or her conclusion. Enlarged photographs may also make useful exhibits for trial. The next step must be to consider how common or rare the wood is. In this regard, it is important to know, or at least be able to look up, something about the uses of a particular wood. Even a relatively rare wood may be common in a certain application. Knowledge of the uses of different woods will help prevent mistakes of this kind.

The final conclusion of a positive wood examination will be that the questioned and known woods are both the same type of wood. Any further interpretation must take into account the factors discussed above. Of course, two woods could not have come from the same piece if they are different genera or species. Only in the case of a physical match will it be possible to say that two pieces of wood could have come from the same piece.

See also: **Paints and Coatings**: Commercial, Domestic and Automotive. **Pattern Evidence**: Tools.

Further Reading

Barefoot AC and Hankins FW (1982) *Identification of Modern and Tertiary Woods*. Oxford: Clarendon Press.

Constantine A (*c*.1959) *Know Your Woods: their identification, properties, and uses from the standpoint of craftsmen, cabinetmakers, carpenters, dealers and students; lumber and veneers; unusual, curious and fabulous woods, and detailed descriptions of 423 individual woods*. New York: Home Craftsman.

Department of Scientific and Instrustrial Research (1953) *An Atlas of End-grain Photomicrographs for the Identification of Hardwoods*. London: HMSO.

Hale JD (1932) *The Identification of Woods Commonly Used in Canada*. Ottawa: Acland.

Ilic J (1987) In: Jeans KF (ed) *The CSIRO Family Key for Hardwood Identification*. Clayton, Australia: Commonwealth Scientific and Industrial Research Organization.

Koehler A (1924) *The Properties and Uses of Wood Prepared in the Extension Division of the University of Wiconsin*. New York: McGraw-Hill.

Panshin AJ and de Zeeuw (*c*.1980) *Textbook of Wood Technology: Structure, Identification, Properties, and Uses of the Commercial Woods of the United States and Canada*. New York: McGraw-Hill.

Parham RA and Gray RL (*c*.1982) *The Practical Identification of Wood Pulp Fibers*. Atlanta, GA: TAPPI Press.

Sharp JB (1990) *Wood Identification: A Manual for the Non-professional*. Knoxville, TN: University of Tennessee, Agricultural Extension Service.

Strelis I and Kennedy RW (1967) *Identification of North American Commercial Pulpwoods and Pulp Fibres*. Toronto: University of Toronto Press.

Plate 1

Plate 1 ACCIDENT INVESTIGATION/Airbag Related Injuries and Deaths This patient sustained a severe corneal abrasion secondary to the membrane forces associated with air bag deployment.

Plate 2

Plate 2 ACCIDENT INVESTIGATION/Driver Versus Passenger in Motor Vehicle Collisions Contact with deploying air bags will result in injuries to the occupants, including abrasions. This patient sustained superficial abrasions overlying the abdomen, secondary to air bag deployment. Such injuries can be matched to the vehicle's air bag.

Plate 3

Plate 3 ACCIDENT INVESTIGATION/Driver Versus Passenger in Motor Vehicle Collisions A degloving injury with underlying fracture. This patient sustained an open fracture, with degloving of the forearm, secondary to contact with the deploying air bag. Matching this injury and the blood transferred to the air bag would assist in identifying the role of this patient.

Plate 4

Plate 4 ACCIDENT INVESTIGATION/Driver Versus Passenger in Motor Vehicle Collisions Examination of the seatbelt webbing and other hardware will be present if the belt sustained sufficient loading. Webbing and hardware should be inspected for evidence of transferred material or imprints from the webbing to the hardware.

Plate 5

Plate 5 ACCIDENT INVESTIGATION/Tachographs The face of a Veeder Root Model 8400 tachograph, *(TVI europe Ltd)*.

Plate 6

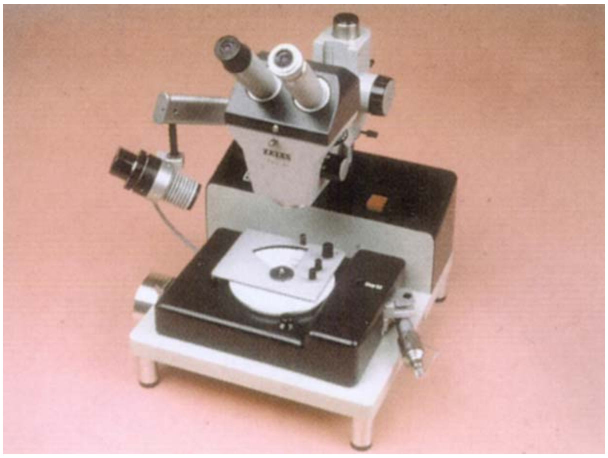


Plate 6 ACCIDENT INVESTIGATION/Tachographs A Kienzle microscope with computer-linking modifications. *(VDO Kienzle (UK) Ltd)*.

Plate 7

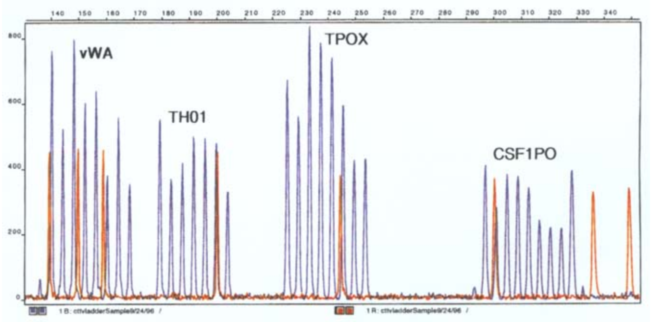


Plate 7 ANALYTICAL TECHNIQUES/Capillary Electrophoresis in Forensic Biology An allelic ladder from the Geneprint cttv multiplex (Promega) consisting of alleles from three genetic loci, CSF1P0, TPOX, THO1, and vWA. The internal standard (labeled in red) is a Genescan 350 ROX (Perkin-Elmer) size standard. Analyzed using an PE/ABI 310 capillary electrophoresis system with laser induced fluorescence detection and a hydroyxethyl cellulose sieving matrix.

Plate 8

Plate 8 ANALYTICAL TECHNIQUES/Separation Techniques Photograph of a SPME fiber inserted into headspace of a sampling vial. (Courtesy R Mindrup, Supelco Inc.)

Plate 9

Plate 9 CAUSES OF DEATH/Traffic Deaths Typical lap/sash passenger distribution of bruising and abrasion.

Plate 10

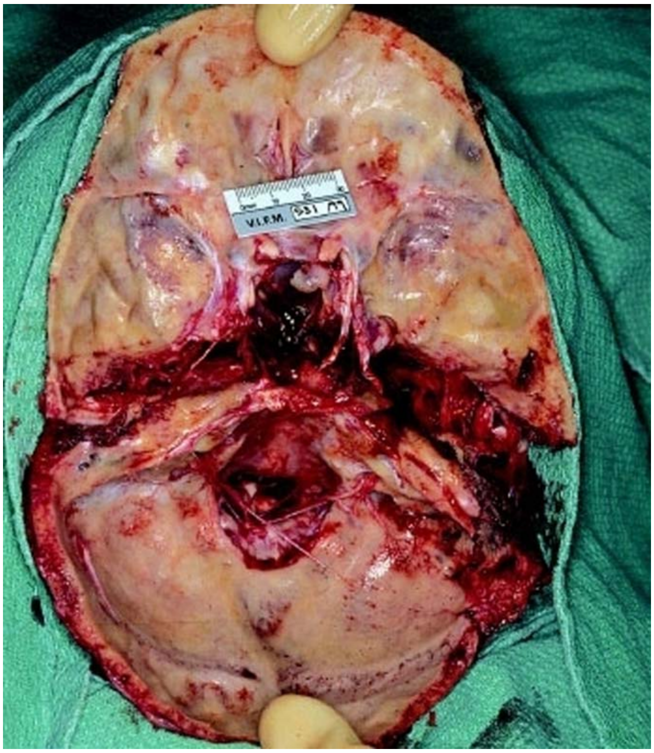


Plate 10 CAUSES OF DEATH/Traffic Deaths Hinge-type base of skull fracture commonly seen as a sequel to blunt lateral head impact.

Plate 11

Plate 11 CAUSES OF DEATH/Traffic Deaths Bruising and abrasion indicating the level of contact with bumper bars.

Plate 12

Plate 12 CAUSES OF DEATH/Traffic Deaths Subcutaneous dissection of the legs to determine accurately the level of impact.

Plate 13

Plate 13 CLINICAL FORENSIC MEDICINE/Evaluation of Gunshot Wounds The contact wound will exhibittriangular shaped tears of the skin. These stellate tears are theresult of injection of hot gases beneath the skin. These gaseswill cause the skin to rip and tear in this characteristic fashion.

Plate 14

Plate 14 CLINICAL FORENSIC MEDICINE/Evaluation of Gunshot Wounds The forensic laboratory can use tests to detect the presence of vaporized lead and nitrates. These tests will assist the forensic investigator in determining the range of fire.

Plate 15

Plate 15 CLINICAL FORENSIC MEDICINE/Recognition of Pattern Injuries in Domestic Violence Victims Blows from a belt can exhibit several different patterns. If the belt impacts on edge, a linear contusion will be produced. If the belt impacts more on its side, a wider contusion may be imprinted. Woven belts will leave a mirror image of the weave imprinted on the skin.

Plate 16

Plate 16 CLINICAL FORENSIC MEDICINE/Recognition of Pattern Injuries in Domestic Violence Victims Contact with the flat portion of a hot iron results in a unique thermal injury. The areas of sparing are from the steam holes.

Plate 17

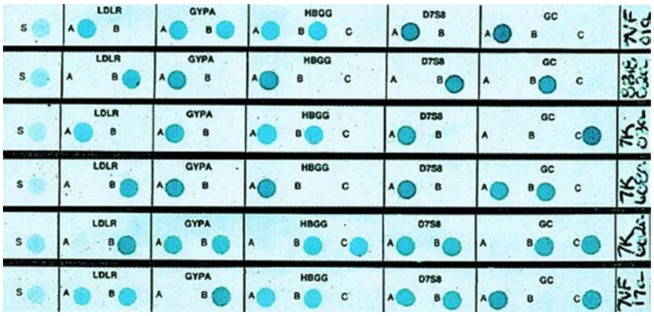


Plate 17 DEOXYRIBONUCLEIC ACID/polymers chain Reaction Examples of Polymarker (PM) strips demonstrating the different types at each locus. Courtesy of PE Biosystems.

Plate 18

Plate 18 DISASTER VICTIM IDENTIFICATION Inadequate body storage facilities. Aftermath of Lauda Air crash at Phu Khao Kao Chan near Bangkok, Thailand, in 1991. Bodies were stored outside with little protection from the heat or insects; decomposition was rapid. Body handlers were inadequately dressed, some with open footwear. Such a mortuary can become a secondary health hazard for the investigators. (Courtesy of Dr Carl KK Leung.)

Plate 19

Plate 19 ENGINEERING The fishing vessel Pescado was recovered from the sea off the south west coast of England to assist the Marine Accident Investigation Branch enquiry into the loss of the vessel and crew. Marine Accident Report 1/98 (1998), London DETR. (Copyright–The Chief Constable, Devon and Cornwall Constabulary.)

Plate 20

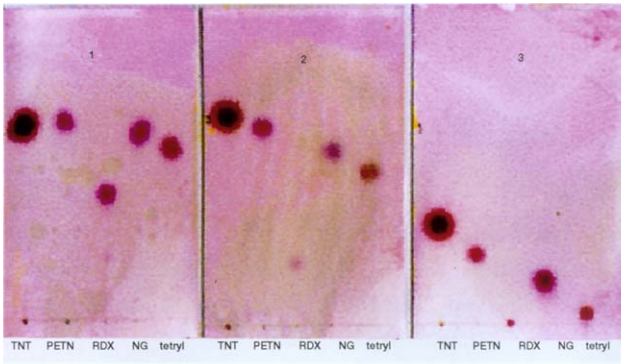


Plate 20 EXPLOSIVES/Analysis Military explosives (TNT, PETN, RDX, NG and tetryl) separated on three TLC solvent systems: (1) 1,2 dichloroethane:acetonitrile (90:10); (2) trichloroethylene:acetone (80:20); (3) petrol ether (b.p. 60-80):ethyl acetate (90:10).

Plate 21

Plate 21 FIBERS/Recovery Recovering fiber debris from a garment using clear adhesive tape.

Plate 22

Plate 22 FIBERS/Recovery Searching for and removing matching fibers from surface debris tapings using stereomicroscopy.

Plate 23

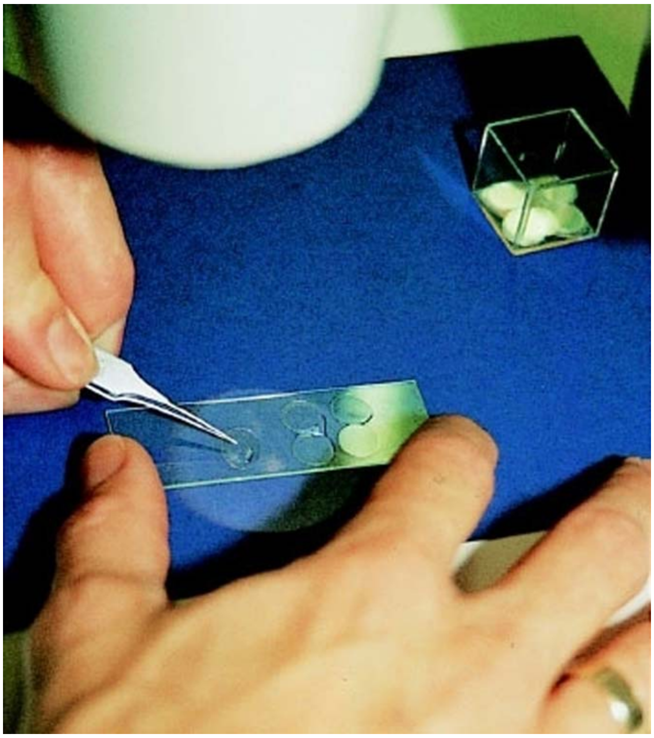


Plate 23 FIBERS/Recovery Mounting recovered fibers onto a microscope slide.

Plate 24

Plate 24 FIBERS/Recovery Seeded comb.

Plate 25

Plate 25 FIBERS/Recovery Searching for recovered fibers on the seeded comb.

Plate 26

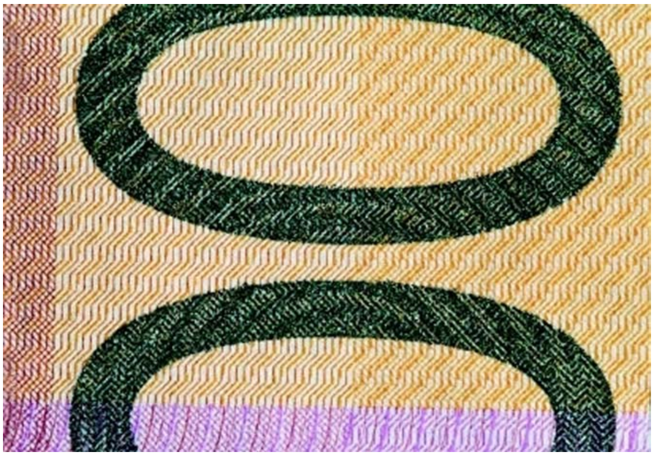


Plate 26 FORGERY AND FRAUD/Counterfeit Currency Screen printing.

Plate 27

Plate 27 FORGERY AND FRAUD/Counterfeit Currency Green-to-black color shifting link on genuine US banknote. Photographed using an angled mirror: green image (below) at normal; black (above) at an oblique angle.

Plate 28

Plate 28 FORGERY AND FRAUD/Counterfeit Currency Fluorescence of genuine US $100 FRN security thread.

Plate 29

Plate 29 FORGERY AND FRAUD/Counterfeit Currency Fluorescence of counterfeit $100 FRN simulated security thread.

Plate 30

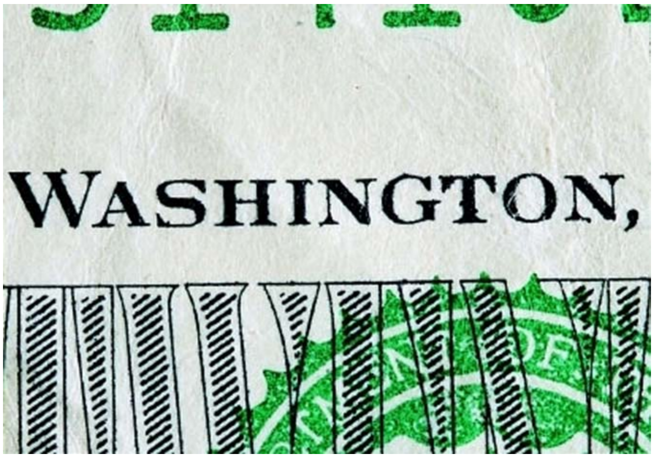


Plate 30 FORGERY AND FRAUD/Counterfeit Currency Genuine US $20 FRN.

Plate 31

Plate 31 FORGERY AND FRAUD/Counterfeit Currency Counterfeit US $20 FRN: full color halftone offset.

Plate 32

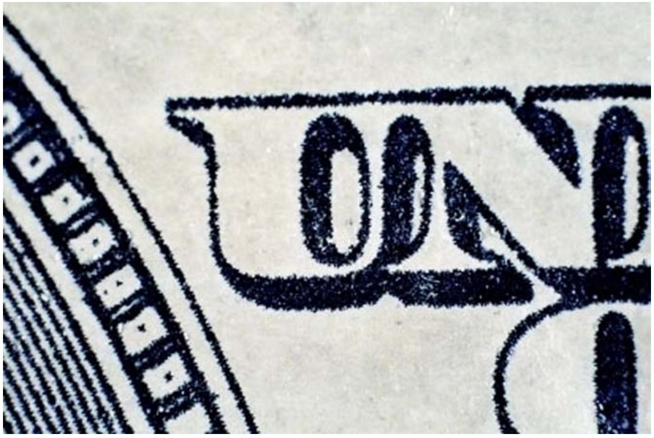


Plate 32 FORGERY AND FRAUD/Counterfeit Currency Counterfeit US $100 FRN: toner full color copier/printer.

Plate 33

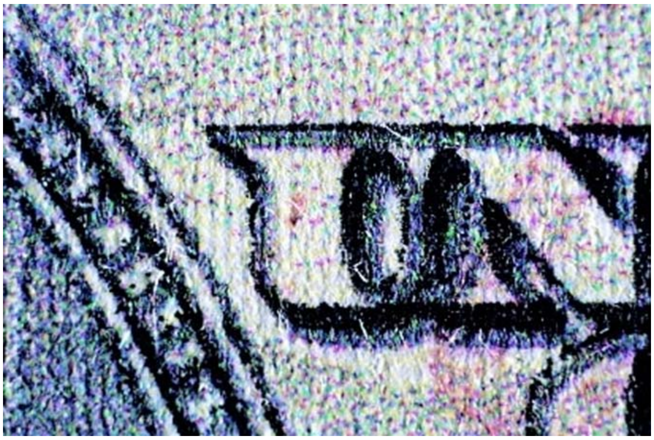


Plate 33 FORGERY AND FRAUD/Counterfeit Currency Counterfeit US $100 FRN: inkjet full color copier/printer.

Plate 34

B

Plate 34 ODONTOLOGY A postmortem shark bite on the remains of someone who had drowned. The arcing lacerations with regular spacing are quite typical of a medium-sized shark bite.

Plate 35

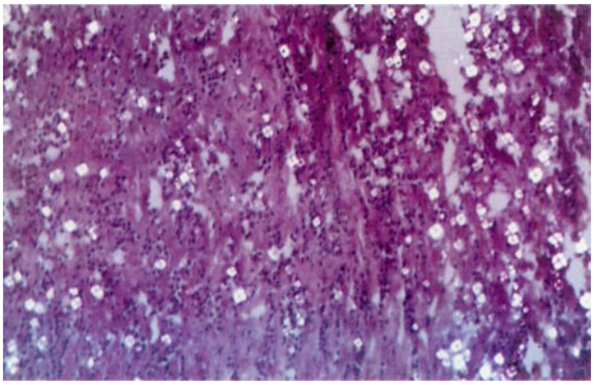


Plate 35 PATHOLOGY/Histopathology Intravenous drug abuse: starch granules and other contaminants at the injection site. H & E; original magnification x 200.

Plate 36

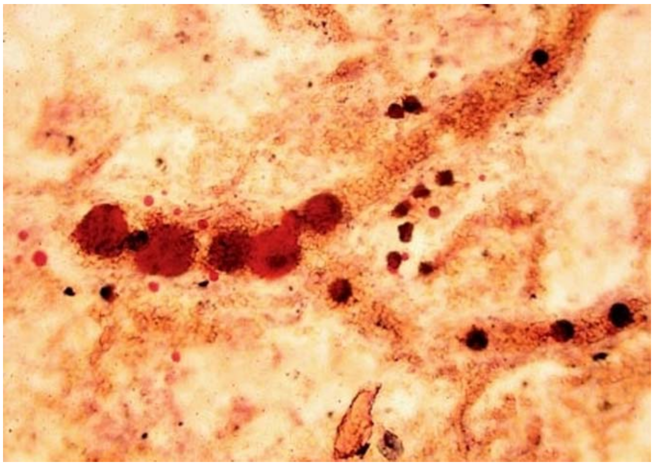


Plate 36 PATHOLOGY/Histopathology Pulmonary fat embolism following multiple fractures. Oil red O; original magnification x 250.

Plate 37

Plate 37 PATHOLOGY/Postmortem Changes hypostasis in a body that has been lying on the back for 12h.

Plate 38

Plate 38 PATHOLOGY/Postmortem Changes Putrefaction: death in a centrally heated flat 4 days previously.

Plate 39

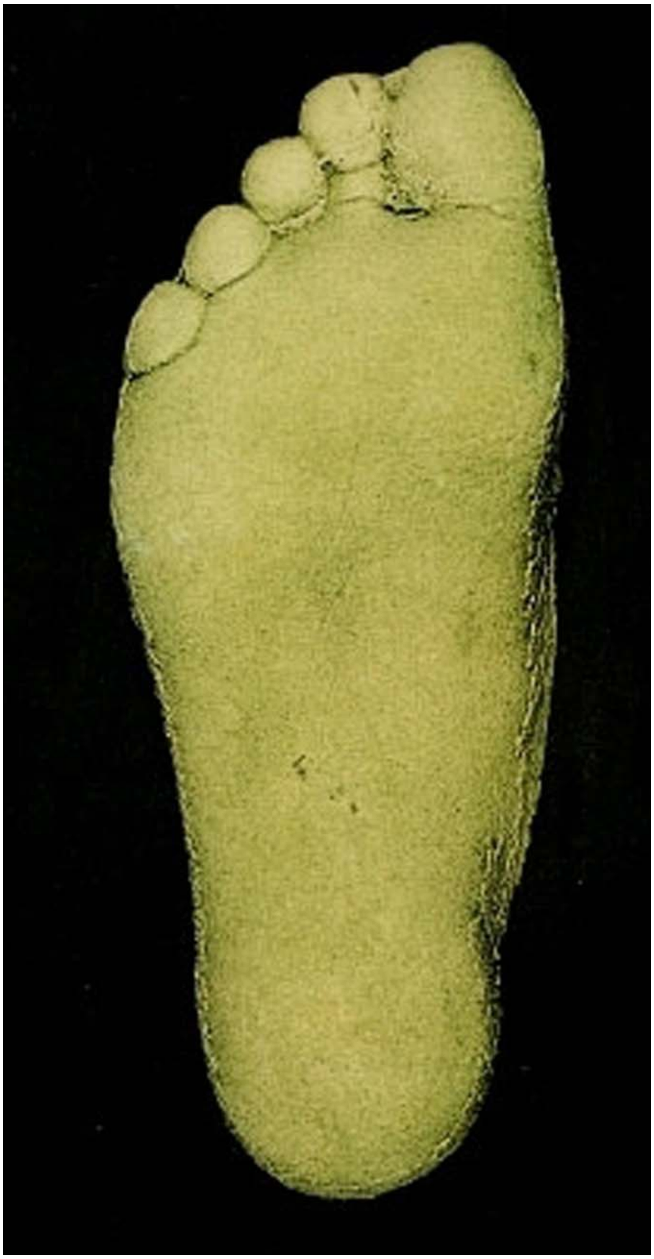


Plate 39 PATTERN EVIDENCE/Bare Footprint Marks Dental stone cast, made from a foam impression material.

Plate 40

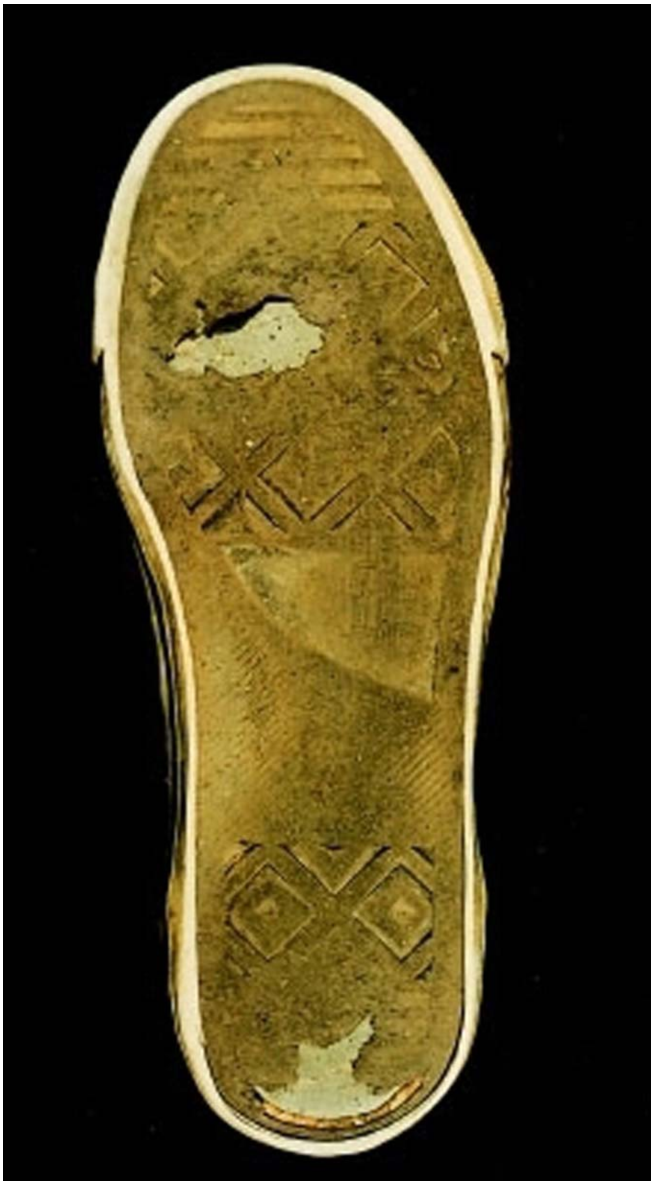


Plate 40 PATTERN EVIDENCE/Bare Footprint Marks Outsole of a running shoe, showing wear areas on the ball of the foot and the heel.

Plate 41

Plate 41 PATTERN EVIDENCE/Bare Footprint Marks Inside top portion of a shoe, showing wear areas on the lining, casued by the top area of the toes.

Plate 42

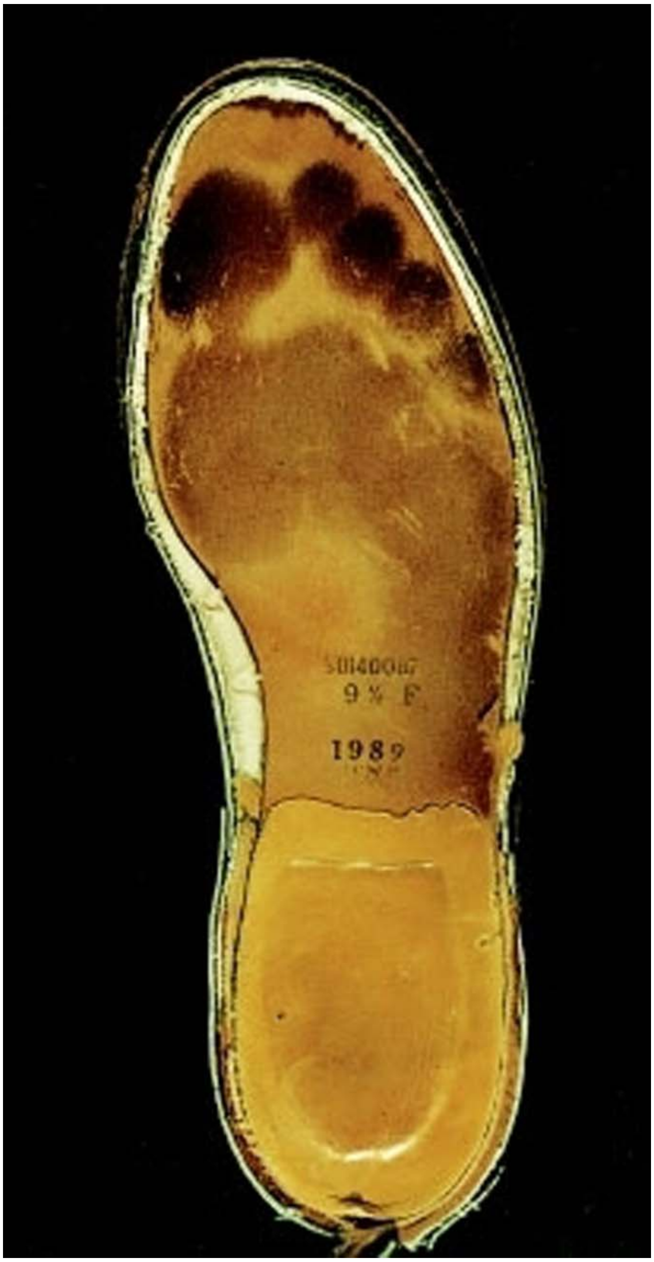


Plate 42 PATTERN EVIDENCE/Bare Footprint Marks Insole of a shoe, showing the darkened and indented sweat areas casued by the weight-bearing areas of a bare foot.

Plate 43

Plate 43 PATTERN EVIDENCE/Footwear Marks Leuco crystal violet is a clear, colorless solution, that can be easily applied at the crime scene or in the laboratory, for the enhancement of bloody footwear impressions. The first picture depicts multiple footwear impressions in blood on a light-colored tile floor. The second picture depicts the same impressions, but after enhancement with leuco crystal violet.

Plate 44

Plate 44 PATTERN EVIDENCE/Tools Macrophotograph showing multicolored plastic insulation material caught in cutter blades of suspect pliers.

Plate 45

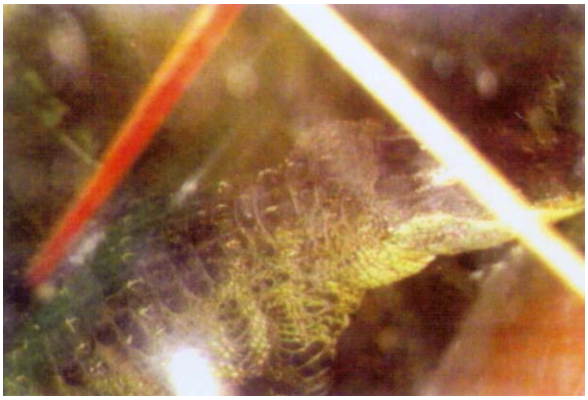


Plate 45 PHOTOGRAPHY AND DIGITAL IMAGING/Overview A ppor image of a crocodile enhanced from a negative, in which we usually recognize its typical textured skin.

Plate 46

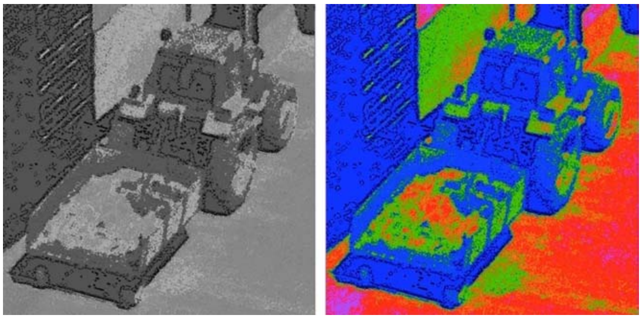


Plate 46 PHOTOGRAPHY AND DIGITAL IMAGING/Digital Imaging Enhancement Color treatment reveals the shapes in the black and white image.

Plate 47

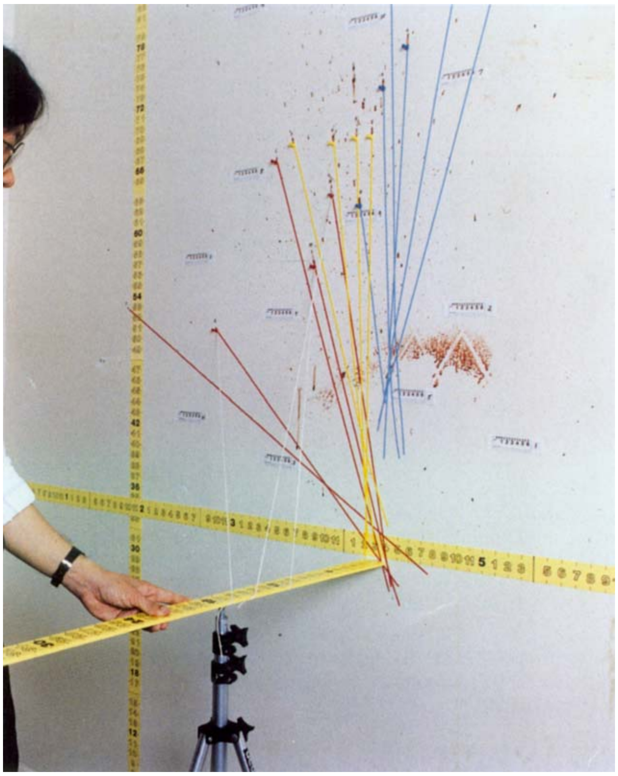


Plate 47 SEROLOGY/Bloodstain Pattern Analysis Bloodstain pattern reconstruction.

Plate 48

Plate 48 WILDLIFE Dr Richard Stroud and Elaine Plaisance examine a North American bear cub to determine cause of death.